Not all wh-dependencies are created equal: processing of multiple wh-questions in Romanian children and adults

Applied Psycholinguistics (2021), 1–40
doi:10.1017/S0142716421000059

ORIGINAL ARTICLE

Not all wh-dependencies are created equal: processing of multiple wh-questions in Romanian children and adults

Anamaria Bentea[1,2,*] and Theodoros Marinis[2,1]

[1]School of Psychology and Clinical Language Sciences, University of Reading, Reading, United Kingdom and [2]Department of Linguistics, University of Konstanz, Konstanz, Germany
*Corresponding author. Email: anamaria.bentea@uni-konstanz.de

(Received 13 December 2019; revised 29 December 2020; accepted 19 January 2021)

Abstract

The aim of this study was to examine the acquisition and processing of multiple *who-* and *which*-questions in Romanian that display ordering constraints and involve exhaustivity. Toward that aim, typically developing Romanian children (mean age 8.3) and adults participated in a self-paced listening experiment that simultaneously investigated online processing and offline comprehension of multiple wh-questions. The study manipulated the type of wh-phrase (*who*/*which*) and the order in which these elements appear (subject–object [SO]/object–subject [OS]). The response to the comprehension question could address the issue of exhaustivity because we measured whether participants used an exhaustive or a non-exhaustive response. Our findings reveal that both children and adults slow down when processing *who*- as compared to *which*-phrases, but only adults show an online sensitivity to ordering constraints in *who*-questions. Accuracy is higher with multiple *who-* than *which*-questions. The latter pose more difficulties for comprehension, particularly in the OS order. We relate this to intervention effects similar to those proposed for single *which*-questions. The lack of intervention effects in terms of reaction times indicates that these effects occur at a later stage, after participants have heard the whole sentence and when they interpret its meaning.

Keywords: multiple wh-questions; language acquisition; sentence processing; self-paced listening; ordering constraints; intervention effects

The syntactic complexity of wh-questions makes them a useful testing ground for the development of grammatical knowledge in children: in order to correctly interpret the subject and object wh-dependencies seen in (1) and (2), the child must (a) have knowledge of movement operations, (b) retain information about the moved elements until they encounter the original position from where movement took place and where the moved elements are assigned an interpretation (this position is called a gap and is marked as an underline below), and (c) establish the corresponding syntactic dependency between the wh-elements and the gap:

1. a. *Who* ____ bit the dog? subject dependency
   b. *Which cat* ____ bit the dog?
2. a. *Who* did the dog bite ____? object dependency
   b. *Which cat* did the dog bite ____?

Crosslinguistic studies on the comprehension of wh-questions, mainly using offline comprehension measures like picture identification and reporting accuracy scores, have found that children already at the age of 4 comprehend subject and object *who*-questions on a par (for English: Avrutin, 2000; Goodluck, 2005, 2010; Hirsch & Hartman, 2006; for French: Bentea & Durrleman, 2013; for Hebrew: Friedmann et al., 2009; for Romanian: Bentea, 2016; a.o). However, there are also studies that report differences in performance between subject and object questions, as children seem to be more accurate with subject than with object questions (for English: Tyack & Ingram, 1977; for Italian: De Vincenzi et al., 1999). Online self-paced-reading studies with adults also show increased processing cost for object extraction manifested in longer reading times for object questions than for subject questions (De Vincenzi, 1991; Meng & Bader, 2000; Schlesewsky et al., 2000; see also Stowe, 1986 for contrasting results).

Moreover, the type of wh-element also affects comprehension as children find object *which*-questions the hardest to interpret correctly. Avrutin (2000) found that English-speaking children (mean age 4.3) give only 48% correct responses for object *which*-questions (2b) compared to an accuracy rate of 80–86% for the other three types of wh-questions, namely subject and object *who*-questions like (1a–b) and subject *which*-questions like (2a). Friedmann et al. (2009) report similar results for Hebrew-speaking children (mean age 4.3) who comprehend object *which*-questions only 58% of the time and subject *which*-questions 78% of the time. Children also performed equally well in the comprehension of subject and object *who*-questions (around 80% correct responses). In a recent study, Contemori et al. (2018) used the visual world paradigm to examine online and offline comprehension of subject and object *which*-questions in English-speaking children aged 5–7. Their offline results are in line with those reported in Avrutin (2000) for English and show a subject–object (SO) asymmetry in children's comprehension of *which*-questions: they were at 63% accuracy for object questions and 95% accuracy for subject questions in a picture matching task. The eye-gaze data in Contemori et al.'s study also indicate that children have more difficulties processing object than subject *which*-questions, and that these difficulties stem from a strong expectation that the first noun encountered will be the subject. In other words, children start with an initial subject preference in object *which*-questions and find it harder to revise and reorient their looks when this initial interpretation turns out to be incorrect.

Children's comprehension difficulties with object *which*-questions have been accounted for in terms of similarity in lexical N restriction between the moved element and the intervening subject (Friedmann et al., 2009; Belletti et al., 2012; Friedmann et al., 2017). This similarity gives rise to intervention effects along the lines of those captured by the principle of Relativized Minimality (RM) in adult grammar (Rizzi, 1990, 2004, 2013, 2018; Starke, 2001). RM states that two elements, X and Y, cannot be connected by movement if Z hierarchically intervenes between them, and Z bears the same morphosyntactic features as X. For example, in order to correctly interpret a sentence like (3), the wh-element *when* (corresponding to the target X) must be related to its trace (Y), but this relation cannot hold because another wh-element *who* (Z) intervenes in the path between *when* and its trace. The violation is triggered by the **identity** relation between the featural specification of the intervener *who* and that of the moved element *when* (in this case, the featural specification of the two elements is +Q[1]):

X Z Y
3. **When* do you wonder ***who*** arrived ___ ?

By extending the application of RM to child grammar, Friedmann et al. (2009) postulate that children encounter difficulties with movement structures in which one element containing a lexical restriction (meaning sequences such as "*the + N*" or "*which + N*") intervenes in the movement of another +N element. According to the featural intervention account, the correct interpretation and production of object *which-N* questions, for example, is hindered by the presence of the subject (*the dog* in [2b] above) because this constituent intervenes between the object *which cat* and the verb *bite* and acts as a competitor in resolving the grammatical dependency between the object and the verb. What makes the subject a potential competitor is the **inclusion** of the +N feature (or lexical restriction) present on the intervening subject in the set of features that also characterizes the moved object. On the other hand, children have no comprehension difficulties with object *who*-questions (2a), as these do not give rise to intervention effects: although the subject *the dog* intervenes between the moved object *who* and its trace, there is a **disjunction** of features between the moved element X and the intervening element Z (the object *who* is specified with a +Q feature, while the intervening subject *the dog* bears a +N feature).

In the context of these findings for the comprehension of single wh-questions, the present study extends the investigation of children's acquisition of wh-dependencies to structures that have received far less attention in language acquisition, namely multiple wh-questions (i.e., questions containing more than one interrogative word). The goals are (a) to examine how children and adults process multiple wh-questions in a language with multiple wh-movement (Romanian) and (b) to uncover the source of difficulty in the comprehension of *who* and *which* multiple wh-questions. Specifically, if children have difficulties with the comprehension of multiple wh-questions, we seek to investigate whether this can be accounted for in terms of featural intervention, that is, in terms of similarity between the features of the moved elements and the features of the intervening constituents.

In the remainder of the introduction, we will present the results of previous studies on the processing and acquisition of multiple wh-questions, as well as the properties of multiple wh-questions in Romanian.

Multiple wh-questions: acquisition and processing

Multiple wh-questions have special syntactic and semantic properties; they involve different dimensions of crosslinguistic variation and therefore lead to additional learning difficulties as compared to single interrogatives illustrated in (1–2) above. The examples in (4) to (6) show that some languages may allow only one wh-phrase to be fronted (English), others may require fronting of all wh-words (Romanian and Slavic languages), whereas some languages do not allow multiple wh-questions at all (Italian, Irish) (see Schulz & Roeper, 2011 for a classification of languages according to availability of multiple wh-movement):

4. *Who* bought *what*? (English)
5. *Cine ce* a cumpărat? (Romanian)
   who what has bought
   "Who bought what?"
6. *Chi ha comprato che cosa? (Italian)
   who has bought what thing
   "Who has bought what?"

Another property of multiple wh-questions is that they may or may not obey Superiority (Chomsky, 1973), a condition that imposes a strict ordering on wh-words and states that the superior, or higher, element must move overtly.[2] So examples like (7a) for English, where the object has moved over the structurally superior subject, violate this constraint:

7. a **What* did *who* buy? (English)
   b. *Which book* did *which child* buy?

Such Superiority effects are cancelled when the wh-expressions are complex phrases of the type *which N*, as illustrated in the example (7b) above (Karttunen, 1977; Pesetsky, 1987; Comorovski, 1989). *Which N* phrases have been termed as D(iscourse)-linked (Pesetsky, 1987, 2000[3]) because these elements, contrary to bare wh-words, prompt an answer chosen from "a set of individuals previously introduced into the discourse" or "from a set that is presumed to be salient to both speaker and hearer" (Pesetsky, 2000, p. 16). Pesetsky (2000) relates the difference in acceptability between (7a) and (7b) to the different movement options available to the two types of wh-phrases. Specifically, in (7b), unlike in (7a), the higher wh-phrase (*which child*) undergoes wh-feature movement, and it is only the lower wh-phrase (*which book*) that undergoes overt phrasal movement, one argument being that *which* or D-linked phrases constitute an exception to the multiple specifier requirement (Pesetsky, 2000: 41) whereby the complementizer in multiple questions requires more than one wh-specifier.[4]

Studies with adults on the processing of multiple wh-questions containing Superiority violations (Hofmeister et al., 2013) argue that questions like (8a) violating Superiority constraints, as the object wh-element is moved by crossing over the higher wh-subject, pose significant processing difficulties that are absent in instances like (8d), which contain the same object–subject order.

8. a. *Mary wondered what who read __
   b. *Mary wondered which book who read __
   c. *Mary wondered what which boy read __
   d. Mary wondered which book which boy read __

Given that *which*-phrases carry additional syntactic and semantic features as compared to bare wh-words, processing of a *which*-phrase involves increased activation and resistance to interference in memory and, as a result, leads to easier retrieval than processing bare wh-words. Indeed, Hofmeister et al. (2013) show that *which*-elements elicit more efficient processing than *who*-constituents in English multiple wh-questions. Sentences with two *which*-phrases receive higher acceptability judgments than minimally different ones with bare wh-words.[5] Acceptability judgments for the *which–who* and *who–which* cases in (8b,c) are intermediate between the *what-who* and the *which-which* ones. Hofmeister et al. (2013) also observe faster reading times at the verb (and its spillover regions) for *which* than for *who* in a self-paced reading task,[6] as well as higher question–answer accuracy for *which–which* than for *who–who* questions.

Therefore, when acquiring multiple wh-questions, children must also be able to distinguish between contexts in which Superiority effects need to be obeyed, as in the case of bare wh-words, and those which represent apparent violations of Superiority, like with *which*-constituents. Apart from acquiring the specific syntactic properties of multiple wh-questions, children also need to learn that this type of question involves pairing relations between the two wh-words (Dayal, 2002; Grohmann, 2003). A question like *Who bought what?* in English requires a pair-list (PL) reading in which the exhaustive sets of possible answers to both *who* and *what* are pairwise linked.[7] The correct answer to (4) above could be "John bought a book and Mary bought a DVD." Two steps are thus necessary to derive a paired list answer: *exhausting* the question domain and *pairing* the two wh-elements. Roeper and de Villiers (2000) note that a paired answer also entails special syntactic relations in syntax, as the subject wh-element must c-command the object or adjunct wh-words (like *where*, *when, how*). When no such c-command relationship holds, as in the case of conjoined questions like "Who ate and what?", paired answers are not required (Krifka, 2001). An investigation into children's answers to multiple wh-questions can therefore inform on the type of structure they assign to these multiple wh-dependencies and reveal (a) whether children have difficulties with exhaustivity, in other words, if they give exhaustive PL answers or they answer only with one pair, or (b) whether children have difficulties with pairing, in other words, if they link one wh-word to another or if they answer only one of the wh-words.

Few studies have looked at the acquisition of multiple wh-questions. In two crosslinguistic studies on the production of multiple interrogatives, Grebenyova (2006, 2011) found that 4- to 6-year-old English-speaking children and children speaking Malayalam had acquired the properties of multiple wh-questions, whereas Russian-speaking children manifested some difficulties with the language-specific syntax of these structures, because they also produced questions with one fronted wh-word and one in-situ, an option that is not allowed in adult Russian. In addition, when exploring the frequency of multiple interrogatives in the input, Grebenyova (2006, 2011) concluded that children's exposure to these structures is minimal.

Roeper and de Villiers (1991) tested the acquisition of PL readings using a question-with-picture task and reported that 4- to 6-year-old English children gave PL answers 78% of the time, while younger children did so in only 32% of the cases. Roeper et al. (2007) also looked at PL readings in multiple wh-questions in English and German and found that German-speaking children acquire such readings at the age of 5 as compared to the age of 6 for English-speaking children. That German children acquire PL reading earlier than English children has been linked to the presence of exhaustivity markers in German, but not in English. The presence of the exhaustivity marker *alles* "all" (Wer kommt *alles*? "Who is coming all?") makes children more likely to give exhaustive answers but also seems to enhance the knowledge that wh-words without *alles* are likely to be exhaustive in German (see Schultz, 2010; Schulz & Roeper, 2011).

Schulz and Roeper's (2011) results on the acquisition of PL readings by German-speaking children with developmental language disorder (DLD, formerly termed as S[pecific] L[anguage] I[mpairement]) showed that these children struggle with providing appropriate PL answers for multiple wh-questions as compared to typically developing children. However, Roeper and Schulz (2011) show that most of the errors in DLD children consisted of providing subject lists (41% of the incorrect answers) and object lists (20%) and that only 16% of the errors consisted of single-pair answers. This suggests that DLD children have more difficulties with pairing, which seems to emerge "independently and later than exhaustivity" (Schulz & Roeper 2011, p. 404).

Gavarró et al (2010) investigated the acquisition of multiple *wh*-questions in 3- to 5-year-old Bulgarian-speaking and Polish-speaking children, through a repetition and a comprehension task. They found that the youngest children correctly repeated multiple interrogatives over 75% of the time and that the number of correct repetitions increased significantly with age. The repetition task also included ill-formed wh-questions with wh-in-situ, Superiority violations, and intervening constituents between the wh-words. The authors report that, in cases of ill-formed questions, children avoided Superiority violations and intervening structures by omitting constituents, and they also raised some in-situ wh-constituents, both in Bulgarian and in Polish. The results of the comprehension task, which included questions with two or three wh-words, show that Bulgarian children give exhaustive answers in 90% of cases even at the age of 3, while at the age of 5, many Polish children still give non-exhaustive answers. The difference reported in Gavarró et al. (2010) between Polish and Bulgarian mirrors the one found for English and German (Roeper et al., 2007): Bulgarian, but not Polish, has plural markers associated with the wh-phrase. This appears to facilitate the early acquisition of exhaustivity in Bulgarian children, whereas Polish children lag behind and provide only 40–60% exhaustive answers.

Bentea (2010) investigated the comprehension of multiple wh-questions by 4- to 6-year-old English, French, and Romanian children and found that English and French children behaved alike in that they gave more PL readings to multiple wh-questions as compared to their Romanian peers. Romanian children, on the other hand, manifested a strong preference to answer only the second wh-element in the structure, although Romanian obligatorily fronts both wh-words. Măniță (2017) also looked at the comprehension of multiple wh-questions with two and three interrogative words in 4- to 6-year-old Romanian children. Her findings show that, even at the age of 6, Romanian children give a low percentage of exhaustive answers to multiple interrogatives; however, they show mastery of exhaustivity in single wh-questions already at the age of 5. This developmental pattern is in line with crosslinguistic findings showing that children recognize exhaustivity in multiple wh-questions later than in single wh-questions or wh-*alles*-questions like in

German (Schulz, 2010; Schulz & Roeper, 2011). Similar to Bentea's (2010) study, the children tested in Măniță (2017) also displayed a preference to answer only one wh-phrase (in this case though, contrary to Bentea, 2010, these answers were mainly to the higher/highest wh-element). Thus, 4- to 6-year-old Romanian children seem to interpret multiple wh-questions as single wh-questions, which can be taken as evidence that they have difficulties interpreting the wh-words as pairs.

In summary, multiple wh-questions present a number of important syntactic and semantic properties that children need to acquire, while faced with impoverished input. In this study, we address a topic that has been understudied in child Romanian and examine how Romanian-speaking children and adults process multiple wh-questions, and to what extent the type of wh-phrase (*who* vs. *which*) can contribute to processing difficulty and overall accuracy in multiple wh-questions.

Properties of Romanian multiple wh-questions

Romanian, like Bulgarian and other Slavic languages, allows for all wh-words to be overtly moved to a clause-initial position. This requirement holds for both bare (9) and D-linked or lexically restricted *which-N* constituents (10).

9. a. Cine pe cine sărută?
who.Nom Acc.who kisses
"Who is kissing whom?"
b. *Pe cine cine sărută?
Acc.who who.Nom kisses
*"Whom is who kissing?"

10. a. Care bunică pe care fată$_j$ o$_j$ sărută?
which.Nom grandmother Acc.which girl$_j$ her$_j$ kisses
"Which grandmother is kissing which girl?"
b. Pe care bunică$_j$ care fată o$_j$ sărută?
Acc.which grandmother$_j$ which.Nom girl her$_j$ kisses
"Which grandmother is which girl kissing?"

11. a. Care bunică pe cine sărută?
which.Nom grandmother Acc.who kisses
"Which grandmother is kissing whom?"
b. Pe care bunică$_j$ cine o$_j$ sărută?
Acc.which grandmother$_j$ who.Nom her$_j$ kisses
"Which grandmother is who kissing?"

Wh-objects in Romanian are marked for Acc by the preposition "pe" and *which*-objects are also obligatorily doubled by a clitic pronoun agreeing in gender and number (*o* "her" in examples [10a,b] and [11b]) (see Dobrovie-Sorin, 1994 for a detailed discussion of the syntax of wh-questions in Romanian). This is an instantiation of the "clitic doubling" phenomenon present in languages like Romanian and Spanish, whereby an accusative or dative clitic pronoun appears together with a co-referential full lexical noun phrase.

While there is a strict ordering among bare wh-elements, since fronting a *who*-object over a *who*-subject is ungrammatical (9b), D-linked or lexically restricted *which*-phrases must always appear clause-initially (10–11), preceding bare phrases, and are known not to show Superiority effects (Comorovski, 1996). This is illustrated by the grammaticality of the *b* examples in (10) and (11), showing that *which*-objects can be fronted across both *which*- and *who*-subjects. From an interpretive standpoint, *which*-expressions are associated with the notion of "givenness" and have been analyzed as topics (see Comorovski, 1996 and Alboiu, 2000). Laezlinger and Soare (2005) and Soare (2009) show that *which*-phrases in Romanian get attracted to a higher position than the position occupied by bare elements at the left periphery of the clause. Given the topic flavor of *which*-constituents, the authors suggest that these phrases target a position that also bears a +Topic feature and which is above the landing site of bare wh-phrases.[8]

Various analyses have been proposed for the order preservation constraint in languages with multiple wh-movement (Richards, 1997; Boškovic, 2002; Krapova & Cinque, 2008; a.o.). The analysis put forth in Krapova and Cinque (2008) traces back the ordering contrast between (9a) and (9b) to a version of RM and featural intervention in which only a whole chain, not just the link of a chain, counts as an intervener and where a chain consists of a moved element and its trace (see also Soare, 2009 and Rizzi, 2017 for an application of this analysis to Romanian multiple wh-questions). The derived structures given in (12a) show that only one occurrence of the object *pe cine*, but not the whole chain, intervenes between *cine* and its trace (represented in angled brackets), and only one occurrence of *cine*, but not the whole chain, intervenes between *pe cine* and its trace. While crossing (or intersecting) chains as in (12a) are allowed in cases of featural identity, as both moved elements bear a +Q feature, nested chains are excluded. In the ungrammatical example (12b), both occurrences of *cine* intervene between *pe cine* and its trace. Under this interpretation, the structure in (12b) is correctly ruled out by RM.

+Q +Q

12. a. Cine pe cine <cine> sărută <pe cine>?

+Q +Q

b. *Pe cine cine <cine> sărută <pe cine>?

If we consider the corresponding sentences, but with lexically restricted *which*-phrases, (13a) is non-problematic because the structure also yields crossing chains. As for (13b), we follow the proposal in Rizzi (2011, 2017) and Villata, Rizzi, and Franck (2016) according to which lexically restricted wh-elements have two available attractors and can target different positions, as compared to bare wh-words which can only be attracted by a +Q feature. Applying this analysis to (13b) results in a configuration where the first *which*-element is attracted by the complex feature conglomerate [+Q, +N], while the second targets a [+Q] position. These structures thus instantiate an inclusion configuration like in the case of single *which* questions, but this inclusion configuration in multiple wh-questions is created by the [+Q] feature. In such cases, even if the whole chain of *care fată* intervenes between the object *pe care bunică* and its trace, the feature of the intervening chain is included in that of the target, so that no violation of RM is triggered. According to this analysis, the same inclusion relation holds in examples like (13c) as well, where the first wh-phrase *pe care bunică* is also attracted by the feature conglomerate [+Q, +N], while the second element, a bare wh-word, can only target a [+Q] position.

+Q +N +Q

13. a. Care bunică pe care fată <care bunică> o sărută <pe care fată>?

+Q +N +Q

b. Pe care bunică care fată <care fată> o sărută <pe care bunică>?

+Q +N +Q

c. Pe care bunică cine <cine> o sărută <pe care bunică>?

The two important points related to the structure of multiple wh-questions in Romanian are (a) that the SO order can be inverted when the object wh-phrase is lexically restricted and (b) that *which*-elements are fronted to a position distinct from and higher than that of bare wh-elements, such that the order preservation constraint, whatever its implementation mechanism, is not operative. Since these questions involve the displacement of two wh-phrases whose featural specification creates either an identity or an inclusion relation, investigating the comprehension of *who* and *which*-multiple questions in Romanian children can shed light on the role of intervention effects in the acquisition of these structures.

The present study

The present study therefore aimed to investigate the acquisition and processing of multiple *who* and *which*-questions in Romanian that display ordering constraints and involve exhaustivity. Specifically, we sought to examine (a) how children and adults process multiple *who*-questions as compared to *which*-questions, (b) whether they display an online sensitivity to ordering constraints in multiple wh-questions, (c) to what extent the type of wh-phrase (*who* vs. *which*) affects overall accuracy in multiple wh-questions, and (d) whether participants provide exhaustive PL answers to multiple wh-questions.

While the other studies that have investigated the comprehension of multiple wh-questions in child Romanian always included an animacy mismatch between wh-words and also contained questions with wh-arguments and wh-adjuncts, in this study we only focus on multiple wh-questions with two wh-arguments and without a mismatch in animacy. In addition, in our study we also manipulate the order of the wh-words, with the wh-subject either preceding or following the wh-object. We adopt a self-paced listening task which offers a segment-by-segment measure of sentence processing, the rationale being that longer listening times at a specific segment in the sentence reflect increased processing difficulties (for children, see Felser et al., 2003; Chondrogianni et al., 2014; Contemori & Marinis, 2014). In addition, because participants must answer the test question verbally, the design not only probes active processing of multiple wh-dependencies based on reaction times (RTs) but also allows us to test whether the online measures correlate with the final interpretation children and adults assign to the sentence.

Method

Participants

Thirty-four Romanian monolingual children between the age of 6 and 9 took part in the study (17 male, 6.11–9.08, $M = 8.3$ years old, $SD = 11$ months). All children were typically developing and had no diagnosed language, hearing, or speech disorders (as reported by the parents). Of the 34 participants, 2 were excluded from further analyses as they did not complete the task. The data of the remaining 32 participants were included in the analyses. Out of these 32 children, there were 14 children aged 6 and 7 ($M = 7.3$ years, $SD = 3$ months), eight 8-year-olds ($M = 8.7$ years, $SD = 4$ months), and ten 9-year-olds ($M = 9.4$ years, $SD = 3$ months). A control group of 20 adults also participated in the study (4 male, $M = 24$; 10 years old, $SD = 52.4$ months). All participants were recruited and tested in Romania. Prior to taking part in the experiment, the children received written parental consent. The adult participants also gave their informed consent. The study was approved by the Ethics Research Committee of the School of Psychology and Clinical Language Sciences, at the University of Reading.

Stimuli

The experiment simultaneously tested online processing (self-paced listening) and offline comprehension (response accuracy) of *who*- and *which*-multiple questions in Romanian children and adults. The participants listened segment-by-segment to embedded questions with two extracted *wh*-phrases (14). As there are no differences in word order or interpretation between embedded and unembedded multiple wh-questions in Romanian, we opted for the use of embedded questions because of the way we set up the task, namely as a game with Paddington the Bear who wants to find out what is happening in various pictures involving princesses and superheroes. More details on the task are provided in the *Procedure* section.

The test sentences varied with respect to the order of the wh-elements (subject-object [SO] vs. object–subject [OS]), as well as the type of wh-words used (only *who*, only *which*, or *which* followed by *who*[9]). Each sentence was preceded by a lead-in introducing the characters in the pictures.

14. *Examples of test sentences used* (the / indicates the segment boundaries; each sentence was divided into 8 segments, *Paddington* being the first segment in each sentence. The lead-in was presented in one block.)

Lead-in:
This is a picture of Jasmine, Elsa, Anna and three grandmothers/
Test sentence:
Paddington/ wants to know/

Subject-Object with two *who*-constituents (SO-*Who*)

a. **cine /pe cine** /sărută /duios /pe obraz /înainte de culcare
who /Acc.who /kisses /lovingly /on the cheek /before bedtime
"who is kissing whom lovingly on the cheek before bedtime."

Object-Subject with two *who*-constituents (OS-*Who*)

b. ***pe cine /cine** /sărută /duios /pe obraz /înainte de culcare
Acc.who /who /kisses /lovingly /on the cheek /before bedtime
"whom who is kissing lovingly on the cheek before bedtime."

Subject-Object with two *which*-constituents (SO-*Which*)

c. **care bunică /pe care prințesă$_j$** /o$_j$ /sărută /pe obraz /înainte de culcare
which grandmother /Acc.which princess$_j$ /her$_j$ /kisses /on the cheek /before bedtime
"which grandmother is kissing which princess on the cheek before bedtime."

Object-Subject with two *which*-constituents (OS-*Which*)

d. **pe care bunică$_j$ /care prințesă** /o$_j$ /sărută /pe obraz /înainte de culcare
Acc.which grandmother$_j$ /which princess /her$_j$ /kisses /on the cheek /before bedtime
"which grandmother which princess is kissing on the cheek before bedtime."

Object-Subject with a *which*-object and a *who*-subject (OS-*WhichWho*)

e. **pe care bunică$_j$ /cine** /o$_j$ /sărută /pe obraz /înainte de culcare
Acc.which grandmother$_j$ /who /her$_j$ /kisses /on the cheek /before bedtime
"which grandmother who is kissing on the cheek before bedtime."

All test sentences contained an equal number of eight segments, the first two segments always being *Paddington/wants to know*. Although the eight segments were of different length, there was a maximum number of three words in each segment. The nouns used matched in gender, number, and animacy, while the verbs used were *chase, cover, follow, hug, kick, kiss, lift, pat, pinch, pull, punch, push, splash*, and *tickle*. The task included a total of 60 items: 10 for every condition in example (14), as well as 10 fillers.[10] The fillers were grammatical subject questions, half of them consisting of embedded single *who*-questions and half of single *which*-questions, illustrated in example (15).

15. *Examples of filler sentences used* (the / indicates the segment boundaries)

a. This is a picture of a woman, a boy, and a man/ Paddington/ wants to know/
cine/ ține/ tortul/ în mână/ seara/ la petrecere
who/ holds/ cake.the/ in hand/ in the evening/ at the party
"who is holding the cake at the party in the evening."

b. This is a picture of two bees and a bird. Paddington/ wants to know/
care albină/ zboară/ rapid/ dimineața/ în grădină/ peste trandafiri
which bee/ flies/ rapidly/ in the morning/ in the garden/ over the roses
"which bee is flying rapidly over the roses in the garden in the morning."

Filler questions contained both transitive (*eat, cut, hold, read*) and intransitive (*sleep, fly*) verbs and were divided as well into eight segments. In using single wh-questions as fillers, we wanted to ensure that participants, especially children, can answer questions correctly and do not have difficulties with this particular type of structure or with the task itself. Moreover, the format of the fillers was similar to that of the test items, in terms of both image display and audio stimuli. Given that both the test sentences and the fillers were introduced by the same two segments *Paddington/ wants to know*, this reduced the type of possible constructions which could follow this introduction.

The test and filler sentences were digitally recorded by a native speaker of Romanian in a soundproof booth. The sentences were segmented using the Audacity software. At the end of each sentence, only one image with three pairs of characters as in Figure 1a–c appeared on the screen, and the participants had to answer the question they had just heard by verbally identifying the correct actions and characters.

After hearing a question with two *who*-constituents, as in (14a) or (14b), participants would see either a picture triad like the one illustrated in Figure 1c containing two actions of the same type and a third different action or a picture triad like in Figure 1a and 1b, in which all the pairs of characters perform the same action, but with reversed Agent–Patient roles. For example, if the participants heard a SO-*Who* question and then saw the picture triad in Figure 1c, they were expected to identify only the pairs that perform the correct action (e.g., *Jasmine is kissing the grandmother and Elsa is kissing the girl.*), while ignoring the irrelevant one (e.g., *cat chasing Anna*). However, we wanted to ensure that participants do not simply rely on verb knowledge when answering these questions and that they can correctly interpret multiple wh-questions as requiring exhaustive PL answers. Therefore, multiple *who*-questions like in (14a) and (14b) were also associated with picture triads as in Figure 1a or 1b. Here the correct answer consisted of exhaustively identifying all three pairs of characters as they all perform the correct action (i.e., someone kissing someone else). Questions containing *which*-constituents (14c–e) were associated with picture triads as in Figure 1a or 1b. In this case, two of the actions in the image corresponded to the correct interpretation of the sentence, while one of the actions corresponded to the reversed Agent–Patient interpretation, in line with stimuli that have been used for testing the comprehension of single *which*-questions. For example, Figure 1a appeared after the SO-*Which* question in (14c). The correct answer in this case would be "*The red grandmother is kissing Anna and the yellow grandmother is kissing Jasmine*" (top right picture and bottom picture), while ignoring the reversed role action in which Elsa is kissing the grandmother in green (top left picture).

Procedure

Participants were tested individually in a quiet room. The self-paced listening task was programmed and ran on a laptop using the PsychoPy software (version 1.90.3, Peirce et al., 2019; Peirce & MacAskill, 2018). Participants had to press the space key to advance from one segment to the next. PsychoPy recorded the time between each key press, which provided the listening times for each segment. The sentences were administered through headphones. All the participants heard all the items, and the order of item presentation was fully randomized automatically with PsychoPy, such

Figure 1. Example of images associated with the different conditions. Each image always depicted three pairs of characters.

that no participant saw the items in the same order. As a result, potential familiarization effects were reduced among items. Although one test sentence appeared both with a SO and an OS order and was associated with an image containing the same pairs of characters, the position of the three pairs was changed between images, as was the direction of the action being performed (left to right or right to left), again to reduce strategic and familiarization effects.

The task was set up as a game about princesses and superheroes, played with Paddington the Bear, in which Paddington wants to find out what is happening in different pictures. The pictures were always about three princesses (Elsa, Anna, Jasmine) and three superheroes (Batman, Superman, Spiderman) interacting with various animals and people. By using pictures that always involved the same characters presented in the introduction, we wanted to reduce the number of new characters that participants had to identify when answering Paddington's questions. Each of these six characters was introduced to the participants when they heard the instructions for the task (see Appendix A) to ensure the participants were familiar with their names. The detailed instructions were then followed by a familiarization phase that included a block of five practice items, which were used to familiarize the participants with the task and with pressing the space bar after each segment. The practice items were constructions similar to the ones used in the experiment, but included a mismatch in animacy as well. The computer screen remained black while participants were listening to the test sentences and it was only when pressing the space bar after the final segment in each sentence that one picture like the ones illustrated in Figure 1 above appeared on the screen. At that moment, the participants had to verbally answer the question and identify the pairs of characters performing the correct action. All the answers were recorded and then transcribed for analysis. The experiment was administered in one session for both children and adults and lasted about 20–30 min.

Predictions

On the basis of the properties of multiple wh-questions and of previous studies on the comprehension of single *who* and *which*-questions (Friedmann et al., 2009; De Vincezi et al. 1999; Bentea, 2016; a.o.), we predict that questions with a SO order should be comprehended better than questions in which the wh-object precedes the wh-subject and that the type of wh-element should also affect comprehension, such that multiple *who*-questions should yield more accurate responses than multiple *which*-questions. Moreover, if children have difficulties computing inclusion configurations in general, even when this inclusion relation is triggered by an overlap in the +Q, but not +N feature, then we should see that they struggle more as compared to adults with the comprehension of structures that instantiate such an inclusion configuration (namely, questions with an OS order like in [13b] and [13c]). Previous studies on the acquisition of multiple wh-questions in Romanian (Bentea, 2010; Măniță, 2017) show that children aged 4–6 give a low percentage of exhaustive responses and that they mainly answer only one of the wh-words. As the children in our experiment are older (6–9), we expect them to give more exhaustive PL answers, in line with crosslinguistic findings that children master exhaustivity in multiple wh-questions around the age of 6.

If asymmetries emerge in online processing as well, then we expect longer RTs at the verb (and possibly its spillover regions) for *which*-questions than for *who*-questions. However, if *which*-phrases pose less processing difficulties as compared to *who*-constituents, based on Hofmeister et al. (2013), then we predict that *which*-items will be processed faster than *who*-items, and that we will also observe faster

RTs at the verb region in *which*-questions, as these elements are easier to retrieve from memory due to increased activation and resistance to interference in memory. If participants show an online sensitivity to ordering constraints, then we expect a slowdown in the order-violating condition upon detecting the ungrammaticality. The pattern is predicted to be qualitatively similar in children and adults, but children might show longer RTs than adults.

Analysis

We analyzed the proportion of accurate responses, as well as the RTs of trials that received correct answers. Accuracy data were analyzed with mixed effects logistic regression (Jaeger, 2008), and RTs were analyzed with mixed effects linear models (Baayen et al., 2008). All analyses were conducted using the *lme4* package (Bates et al., 2015) in R (R Core Team, 2019), and figures were produced using the package *ggplot2* (Wickham, 2016).

Accuracy

We used generalised linear mixed effects regression (GLMER) modeling to analyze the accuracy data in R. The statistical analysis was performed in two stages. The first analysis focused on the comparison between the SO and OS orders in the *who* and *which* conditions for both children and adults. The model included WhType (*Who* vs. *Which*), WhOrder (SO vs. OS), and Group (Children vs. Adults), as well as their interaction as fixed predictors. The second analysis considered the OS conditions separately to evaluate whether the comprehension of multiple questions with an OS word order was modulated by the effect of WhType (*Who* vs. *WhichWho* vs. *Which*) and whether this effect differed across the two Groups (Adults vs. Children). All the fixed factors in the two analyses were coded using sliding difference or repeated contrasts which test consecutive factor levels against each other (Schad et al., 2019). The contrasts were coded as follows – WhType: *Who* (-1) versus *Which* (1); WhOrder: SO (-1) versus OS (1); Group: Adults (-1) versus Children (1). In the model with OS structures only, we specified the following contrast matrix for WhType (c2 vs. 1, with *Who* coded as −1, *WhichWho* as 1, and *Which* as 0; c3 vs. 2, with *Who* coded as 0, *WhichWho* as −1 and *Which* as 1). For the random effect structure of the models, we followed current guidelines in psycholinguistics and chose parsimonious mixed models (Bates et al., 2018), because these models are more suited for analyzing the typical samples included in psycholinguistic research (Matuschek et al., 2017). The first model included participant and items as random intercepts and by-participant random slopes for WhType; the second model only included participant and items as random effects. The goodness-of-fit of alternative models for the random effects structure was assessed by comparing the Akaike information criterion scores (Akaike, 1974). A decrease of at least 2 in the Akaike information criterion scores means that the inclusion of a factor significantly improves the goodness-of-fit of the model. We also specified the bobyqa optimizer in the glmer function in order to sustain model convergence.

Reaction times

The RT data were analyzed with a linear mixed effects regression (LMER) model in R. The RTs for each segment were analyzed separately and the analysis was performed again in two steps, as for the accuracy data. In a first instance, we compared the RTs for conditions (a) to (d) in the example (14) above, so multiple *who* and multiple *which* questions with a SO and OS order. Group (Children vs. Adults), WhType (*Who* vs. *Which*), and WhOrder (SO vs. OS), as well as their interaction, were specified as fixed predictors. Second, we examined the effect of WhType (*Who* vs. *WhichWho* vs. *Which*) and Group (Children vs. Adults) and their interaction on the RTs for the OS conditions only. Like in the case of the analysis for accuracy, we coded the fixed predictors using repeated contrasts and using the same contrast matrix. The final maximal models supported by the data included participant and items as random intercepts and by-participant random slopes for WhType. The goodness-of-fit of alternative models for the random effects structure was assessed by comparing the Akaike information criterion scores (Akaike, 1974). p-Values were calculated by Satterthwaite's approximation for denominator degrees of freedom, using the *lmerTest* package (Kuznetsova et al., 2017).

Results

The accuracy scores inform us about the final interpretation that participants assign to multiple wh-questions. RTs, that is, the listening times at each segment in the question, inform us on how sentences are processed incrementally as they unfold and at which segment(s) in the sentence participants encounter processing difficulties. Figures and averages are shown in untransformed measures for ease of interpretation, but statistical analyses were performed on log-transformed measures.

Accuracy

Table 1 indicates accuracy scores for both children and adults in each experimental condition. For each test trial, an answer was coded as accurate when the participants identified in the image all the correct actions corresponding to the question, as the correct interpretation of multiple wh-questions in Romanian requires a PL answer. The correct answer for the fillers required identifying only one out of three characters. The fact that the accuracy rate for the filler trials was very high in both children and adults shows that participants do not have difficulties with the comprehension of single embedded questions. While adults were at ceiling, children's accuracy rate for the filler trials was 95% (96% response accuracy for subject *who*-questions and 94% response accuracy for subject *which*-questions).

For the analysis, we will first focus on the comparison between multiple *who*-questions and multiple *which*-questions with a SO and OS order of constituents. Figure 2 illustrates the distribution of accurate responses (in percentages) in the *who* and *which* experimental conditions for both children and adults. The results indicate that both children and adults (a) comprehend multiple *who*-questions better than multiple *which*-questions and (b) that there is a difference in accuracy between the SO and the OS conditions in *which*, but not in *who*-questions.

Table 1. Number and percentage of correct responses and standard deviation per condition for children and adults

Condition		Children	Adults
SO-*Who*	Number	277/319[13]	193/200
	Total (%)	87	97
	SD	0.31	0.18
OS-*Who*	Number	285/319	193/200
	Total (%)	89	97
	SD	0.27	0.18
SO-*Which*	Number	196/320	185/200
	Total (%)	61	93
	SD	0.49	0.24
OS-*Which*	Number	164/320	143/200
	Total (%)	51	72
	SD	0.50	0.45
OS-*WhichWho*	Number	220/320	189/200
	Total (%)	69	95
	SD	0.47	0.23

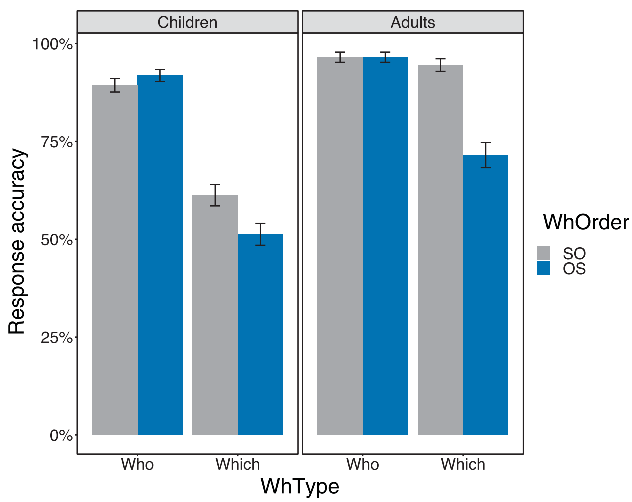

Figure 2. Overall accuracy for *Who* and *Which* multiple wh-questions with a SO and OS order (SO = subject–object, OS = object–subject). The bars represent the standard error to the mean.

Table 2. Accuracy Model 1 output (formula: glmer *(Accuracy ˜ WhType + WhOrder + Group + WhType:WhOrder + WhType:Group + WhOrder:Group + (1 + WhType | ID) + (1 | Item), family = binomial, control = glmerControl(optimizer = "bobyqa"), data)*)

Fixed effects	Estimate	*SE*	*z*-value	*p*-value
(Intercept)	2.495	0.238	10.452	<.001
WhType: Who versus Which	−2.609	0.363	−7.171	<.001
WhOrder: SO versus OS	−0.590	0.191	−3.084	<.01
Group: Adults versus Children	−1.220	0.415	−2.934	<.01
$\text{WhType}_{\text{Who versus Which}} \times \text{WhOrder}_{\text{SO versus OS}}$	−1.247	0.339	−3.674	<.001
$\text{WhType}_{\text{Who versus Which}} \times \text{Group}_{\text{Adults versus Children}}$	−1.228	0.580	−2.118	.034
$\text{WhOrder}_{\text{SO versus OS}} \times \text{Group}_{\text{Adults versus Children}}$	1.343	0.336	3.992	<.001

Table 2 gives the output for the fixed effects of the final GLMER fit to children's and adults' accuracy scores for SO and OS multiple wh-questions in the conditions with two *who* and two *which* constituents.

The model revealed a significant main effect of WhType, showing that multiple *which*-questions ($M = 67\%$) are significantly less accurate than multiple *who*-questions ($M = 93\%$). The effect of WhOrder was also significant and indicates that multiple wh-questions with an OS order ($M = 76\%$) lead to lower response accuracy than multiple wh-questions with a SO order ($M = 83\%$). The significant effect of Group reveals that children perform significantly less accurately ($M = 73\%$) than adults ($M = 90\%$). In order to explain the direction of the statistically significant interactions in the final model, we nested the pairwise comparisons. The significant interaction between WhType and WhOrder and subsequent pairwise comparisons show that there is no significant difference between multiple SO-*Who* and multiple OS-*Who* questions ($\beta = -0.033$, $SE = 0.310$, $z = -0.108$, $p = .914$), while multiple SO-*Which* questions were overall more accurate than multiple OS-*Which* questions ($\beta = 1.214$, $SE = 0.186$, $z = 6.520$, $p < .001$). A significant interaction between WhType and Group and subsequent pairwise comparisons reveal that there is no significant difference in performance between children and adults in the case of multiple *who*-questions ($\beta = -0.606$, $SE = 0.635$, $z = -0.955$, $p = .339$), but that children perform significantly less accurately than adults with multiple *which*-questions ($\beta = -1.835$, $SE = 0.334$, $z = -5.494$, $p < .001$). The significant interaction between WhOrder and Group and subsequent pairwise comparisons indicate that multiple wh-questions with a SO order yield significantly lower accuracy in children as compared to adults ($\beta = -1.892$, $SE = 0.467$, $z = -4.051$, $p < .001$), while performance with OS multiple wh-questions does not differ significantly between the two groups ($\beta = -0.548$, $SE = 0.429$, $z = -1.277$, $p = .201$). We also constructed an additional model to test for the interaction between WhType, WhOrder, and Group, and the results of this model were compared to the model with two-way interactions only by means of the *anova* function. This revealed no significant difference between the two models, based on the *p*-value associated with the chi-square-distributed likelihood ratio ($p = .120$), and thus no significant three-way interaction between these three factors.

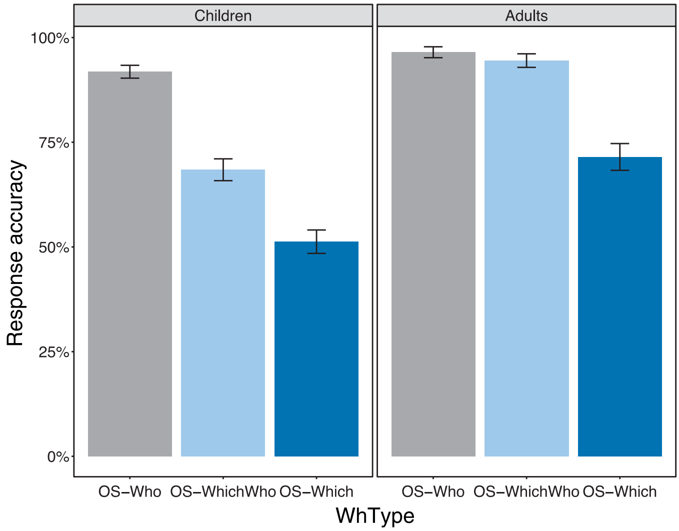

Figure 3. Distribution of correct responses (in percentages) in OS multiple wh-questions with two *who*-constituents (OS-*Who*), a *which*-object and a *who*-subject (OS-*WhichWho*), and two *which*-constituents (OS-*Which*). The bars represent the standard error to the mean.

Table 3. Accuracy Model 2 output (formula: glmer *(Accuracy ˜ WhType + Group + WhType:Group + (1 | ID) + (1 | Item), family = binomial, control = glmerControl(optimizer = "bobyqa"), data)*)

Fixed effects	Estimate	*SE*	*z*-value	*p*-value
(Intercept)	2.176	0.247	8.786	<.001
WhType: OS-*Who* versus OS-*WhichWho*	−1.142	0.459	−2.489	.012
WhType: OS-*WhichWho* versus OS-*Which*	−1.7550	0.418	−4.193	<.001
Group: Adults versus Children	−1.557	0.389	−3.997	<.001
$\text{WhType}_{\text{OS-}\textit{Who}\text{ versus OS-}\textit{WhichWho}} \times \text{Group}_{\text{Adults versus Children}}$	−1.775	0.564	−3.143	<.01
$\text{WhType}_{\text{OS-}\textit{WhichWho}\text{ versus OS-}\textit{Which}} \times \text{Group}_{\text{Adults versus Children}}$	1.437	0.420	3.420	<.001

Moving on to examine the results obtained for multiple wh-questions with an OS order, we observe from Figure 3 that the OS-*Which* questions yielded the lowest accuracy scores in both children and adults. Whereas adults comprehended OS-*Who* and OS-*WhichWho* questions equally well, children's performance with OS-*WhichWho* questions was less accurate than their performance with OS-*Who* questions.

The significant main effect of WhType in the analysis for multiple wh-questions with an OS order (Table 3) reveals that OS-*Who* questions ($M = 94\%$) are significantly more accurate than OS-*WhichWho* questions ($M = 78\%$), and that the OS-*WhichWho* conditions are comprehended significantly better than the

OS-*Which* conditions ($M = 59\%$). The significant main effect of Group indicates that, when we also consider the results for OS-*WhichWho* questions, children are overall less accurate ($M = 70\%$) with OS multiple wh-questions than adults ($M = 86\%$). The significant interaction between OS-*Who* versus OS-*Which* and Group and follow-up pairwise comparisons show that, for children, OS-*Who* questions are significantly more accurate than OS-*WhichWho* questions ($\beta = 2.030$, $SE = 0.451$, $z = 4.501$, $p < .001$), but no significant difference emerges between the two conditions in adults ($\beta = -0.255$, $SE = 0.614$, $z = 0.414$, $p = .909$). Moreover, the significant interaction between OS-*WhichWho* versus OS-*Which* and Group and subsequent pairwise comparisons indicate that for children, and even more so for adults, OS-*WhichWho* questions yield significantly higher accuracy than OS-*Which* questions (children: $\beta = 1.036$, $SE = 0.407$, $z = 2.547$, $p = .029$; adults: $\beta = 2.474$, $SE = 0.523$, $z = 4.733$, $p < .001$).

Given that the child group covers a large age range (6-9 years old), we ran further analyses on the child data only and included age in months as a continuous variable in the model in order to test whether age plays a role in modulating the comprehension of multiple wh-questions in children. The analysis revealed a significant effect of Age ($\beta = 0.054$, $SE = 0.022$, $z = 2.447$, $p < .01$), showing that older children give overall more accurate responses than younger children. The interaction between WhOrder and Age was also significant ($\beta = 0.030$, $SE = 0.014$, $z = 2.132$, $p < .05$), indicating that the effect associated with the order of the wh-elements increases with age. The results of a post-hoc analysis (see Appendix B for the full output of each model), aiming to disentangle the effects of Age and WhOrder on multiple *who*-questions and multiple *which*-questions, reveal a significant effect of Age in both multiple *who*-questions ($\beta = 0.098$, $SE = 0.039$, $z = 2.466$, $p < .05$) and multiple *which*-questions ($\beta = 0.036$, $SE = 0.012$, $z = 2.949$, $p < .01$). In other words, older children tended to perform better on the task than younger children. However, WhOrder was a significant predictor only in the model that considered multiple *which*-questions ($\beta = 0.471$, $SE = 0.171$, $z = 2.741$, $p < .01$). Specifically, multiple *which*-questions containing a SO order yielded more accurate responses than multiple *which*-questions with an OS order. For multiple *which*-questions, we also found a significant interaction between WhOrder and Age ($\beta = 0.034$, $SE = 0.015$, $z = 2.235$, $p < .05$). Subsequent pairwise comparisons show that older children comprehend SO-*Which* questions containing a SO order significantly better than younger children ($\beta = 0.054$, $SE = 0.015$, $z = 3.612$, $p < .001$). Although response accuracy for OS-*Which* questions also increases with age ($\beta = 0.019$, $SE = 0.014$, $z = 1.350$, $p = .177$), this does not reach significance.

An asymmetry between multiple questions with two *who*-elements and those containing two *which*-elements also surfaces when analyzing the errors that children and adults make in answering these questions. The three main types of wrong answers or errors, summarized in Table 4, are (a) over-exhaustive answers (when participants identify all the pairs in the image, even though one of them does not match the action), (b) singleton answers (when participants answer only one wh-word, either the wh-subject or the wh-object, and they exhaustively list all the individuals involved in the corresponding action), and (c) role reversals (when participants reverse the Agent–Patient roles). Children not only make more errors than adults but also make a significant number of singleton errors, that is, they answer only one of the wh-words. While adults very rarely give such answers (4 out of a total of 97 errors), this is the most

Table 4. Type of errors (percentages and raw numbers out of total number of errors per condition) in children and adults for both SO and OS multiple wh-questions with two *who*-constituents (*Who*), two *which*-constituents (*Which*), and a *which*-object and a *who*-subject (*WhichWho*)

Group	Type of errors	*Who*	*Which*	*WhichWho*
Children	Over-exhaustive	21 (16/76)	11.4 (32/280)	22 (22/100)
	Singleton	73.7 (56/76)	22.2 (63/280)	34 (34/100)
	Wrong action	5.3 (4/76)	0 (0/280)	0 (0/100)
	Reversed role	0 (0/76)	66.4 (185/280)	44 (44/100)
Adults	Over-exhaustive	100 (14/14)	8.3 (6/72)	18.2 (2/11)
	Singleton	0 (0/14)	2.8 (2/72)	18.2 (2/11)
	Reversed role	0 (0/14)	88.9 (64/72)	63.6 (7/11)

Table 5. Type and number of errors for each condition in children by age group

Age Group	Error Types	*Who*	*Which*	*WhichWho*
6-7 years	Over-exhaustive	12/76	25/280	18/100
	Singleton	49/76	45/280	27/100
	Wrong action	1/76	0/280	0/100
	Reversed ole	0/76	76/280	17/100
8 years	Over-exhaustive	0/76	3/280	2/100
	Singleton	3/76	9/280	2/100
	Wrong action	1/76	0/280	0/100
	Reversed role	0/76	50/280	14/100
9 years	Over-exhaustive	4/76	4/280	2/100
	Singleton	4/76	9/280	5/100
	Wrong action	2/76	0/280	0/100
	Reversed role	0/14	59/280	13/100

common type of error that children make when answering questions with two *who*-elements and the second most common type of error in questions with two *which*-phrases and in *WhichWho* questions. An analysis of children's errors by age group (Table 5) shows that the younger children (6- to 7-year-olds) give more singleton responses consisting of exhaustive subject or object lists than the older children. From the total of 152 such errors, 121 appear in the 6- to 7-year-old group, compared to only 14 for the 8-year-old group and 17 for the 9-year-old group. The most frequent error that both children and adults make in multiple *which*-questions is role reversal, indicating that both groups have more difficulties mapping the correct argument role unto the two wh-words when both are lexically restricted.

To summarize, the accuracy data show that children made significantly more errors than adults in their comprehension of multiple *which*-questions. Furthermore, children also showed lower accuracy in the comprehension of multiple wh-questions with an OS order, both when the questions contained two *which*-phrases and when the questions contained a *which*-object phrase and a *who*-subject. No differences emerged with respect to the comprehension of multiple *who*-questions.

Reaction times

The RT analyses were performed on residual RTs for accurate trials only. These were calculated by subtracting the participants' raw RTs from the total duration of each segment. Residual RTs were further screened for extreme values and outliers (see Marinis, 2010). Extreme values were defined as RTs below –1000 ms and above 2500 ms on the basis of histograms and were eliminated from the dataset. Outliers were defined as RTs above and below 3 standard deviations for each condition separately per participant and item and were replaced with the mean RT for each condition per participant and item. Extreme values and outliers comprised 1.6% of the data (2.3% of the data for children and 0.8% of the data for adults). Although each test sentence started with *Paddington / wants to know /*, we only analyzed the RTs for six segments, Segment 1 being the first wh-word. Figures 4-7 show the RTs in milliseconds for children and adults at each of the segments of interest, starting from the first wh-word that the participants hear. Like for the accuracy data, we first analyze the RTs for multiple *who*-questions and multiple *which*-questions with a SO and OS order, which allows us to examine how multiple wh-questions are processed in real-time and whether differences appear between *who* and *which*-phrases. We only report the significant effects and interactions in the text and we present the data segment-by-segment. Figure 4 reports the RTs for children and Figure 5 shows the RTs for adults.

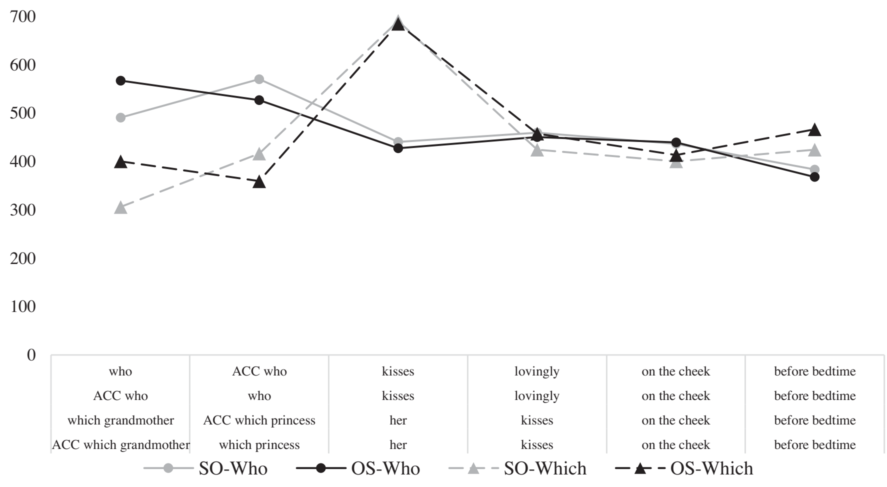

Figure 4. Distribution of children's RTs (in ms) for the different sentence segments (English translation) in four experimental conditions (SO-*Who*, OS-*Who*, SO-*Which*, OS-*Which*).

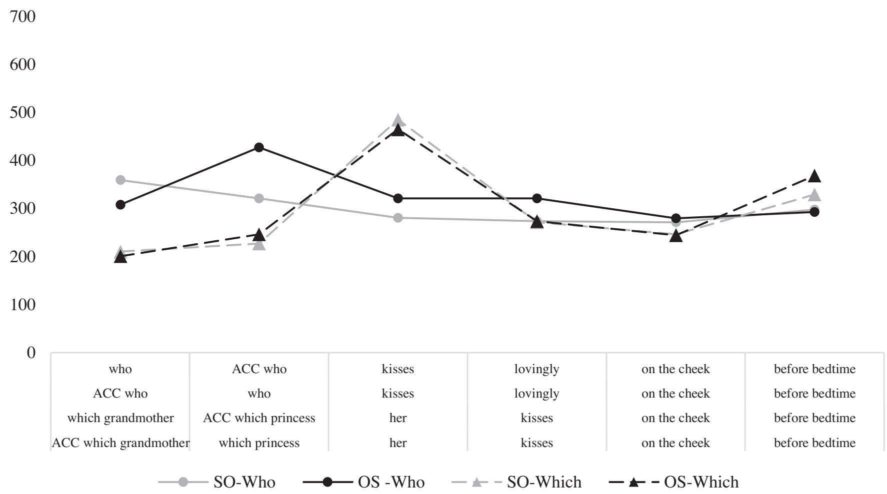

Figure 5. Distribution of adults' RTs (in ms) for the different sentence segments (English translation) in four experimental conditions (SO-*Who*, OS-*Who*, SO-*Which*, OS-*Which*).

Segment 1

At Segment 1, there was (a) a significant main effect of Group ($\beta_{\text{Adults vs. Children}} = 0.203$, $SE = .091$, $t = 2.217$, $p = .031$), reflecting longer RTs in children as compared to adults; (b) a main effect of WhType ($\beta_{\text{Who vs. Which}} = -0.325$, $SE = 0.050$, $t = -6.440$, $p < .001$), indicating significantly shorter RTs for *which*-phrases as compared to *who*-phrases; and (c) a main effect of WhOrder ($\beta_{\text{SO vs. OS}} = 0.044$, $SE = 0.019$, $t = 2.348$, $p = .019$), which shows that wh-objects yield significantly longer RTs than wh-subjects. Moreover, we found a significant interaction between WhOrder and Group, indicating that the order of the wh-elements affected children and adults in a different way ($\beta_{\text{SO vs. OS} \times \text{Adults vs. Children}} = 0.202$, $SE = 0.037$, $t = 5.393$, $p < .001$). Subsequent pairwise comparisons reflect that the OS conditions yield significantly longer RTs in children as compared to adults ($\beta = 0.304$, $SE = 0.093$, $t = 3.248$, $p < .01$), while there is no significant difference in children's and adults' RTs in the SO conditions ($\beta = 0.102$, $SE = 0.093$, $t = 1.093$, $p = .279$). We also found a significant interaction between WhOrder and WhType at Segment 1 ($\beta_{\text{SO vs. OS} \times \text{Who vs. Which}} = 0.080$, $SE = 0.037$, $t = 2.129$, $p = .033$). Subsequent pairwise comparisons at Segment 1 showed that the SO order yielded significantly shorter RTs than the OS order in the case of *which*-questions ($\beta = -0.084$, $SE = 0.029$, $t = -2.922$, $p < .01$). The effect of WhOrder in *who*-questions goes in the same direction, namely SO shorter than OS, but it is not significant ($\beta = -0.004$, $SE = 0.024$, $t = -0.178$, $p = .858$).

Segment 2

Like in Segment 1, we found a significant main effect of Group ($\beta_{\text{Adults vs. Children}} = 0.235$, $SE = 0.096$, $t = 2.431$, $p = .018$) and of WhType ($\beta_{\text{Who vs. Which}} = -0.351$, $SE = 0.052$, $t = -6.643$, $p < .001$). The analysis revealed as well a significant interaction between

WhOrder and Group ($\beta_{\text{SO vs. OS} \times \text{Adults vs. Children}} = -0.286$, $SE = 0.041$, $t = -6.842$, $p < .001$), reflecting that adults show significantly shorter RTs for questions with a SO order than for those with an OS order ($\beta = -0.155$, $SE = 0.031$, $t = -4.920$, $p < .001$), whereas children show the reverse pattern, as their RTs are significantly longer in the SO order compared to the OS order ($\beta = 0.132$, $SE = 0.028$, $t = 4.393$, $p < .001$). The interaction between WhOrder and WhType in Segment 2 was marginally significant ($\beta_{\text{SO vs. OS} \times \text{Who vs. Which}} = -0.077$, $SE = 0.042$, $t = -1.847$, $p = .064$), and subsequent pairwise comparisons reveal that the difference between the SO conditions and the OS conditions in the case of *who*-questions approached significance, with SO faster than OS ($\beta = -0.050$, $SE = 0.027$, $t = -1.854$, $p = .063$), but no such difference was found in the case of *which*-questions ($\beta = 0.027$, $SE = 0.032$, $t = 0.846$, $p = .397$).

Segment 3

A main effect of Group also emerged in Segment 3 ($\beta_{\text{Adults vs. Children}} = 1.996$, $SE = 7.312$, $t = 2.799$, $p < .01$), as children have longer RTs than adults, together with a main effect of WhType ($\beta_{\text{Who vs. Which}} = 3.568$, $SE = 5.672$, $t = 6.290$, $p < .001$), which reflects significantly longer RTs at the clitic region in questions with two *which* elements compared to RTs at the verb region in questions with two *who* phrases. The significant interaction between WhOrder and Group at Segment 3 ($\beta_{\text{SO vs. OS} \times \text{Adults vs. Children}} = -6.764$, $SE = 3.218$, $t = -2.102$, $p = .035$) and subsequent pairwise comparisons indicate that the SO condition was significantly longer than the OS in children ($\beta = 0.048$, $SE = 0.021$, $t = 2.270$, $p = .023$), but there was no difference in RTs between the SO and OS orders for adults ($\beta = -0.032$, $SE = 0.023$, $t = -1.357$, $p = .175$). In addition, we found a significant interaction between WhOrder, WhType, and Group in Segment 3 ($\beta_{\text{SO vs. OS} \times \text{Who vs. Which} \times \text{Adults vs. Children}} = 1.735$, $SE = 6.434$, $t = 2.696$, $p < .01$). The pairwise comparisons reveal significant differences in RTs in the following pairs: SO-*Who* shorter than OS-*Who* for adults ($\beta = -0.081$, $SE = 0.032$, $t = -2.532$, $p = .011$) and SO-*Who* longer than OS-*Who* for children ($\beta = 0.072$, $SE = 0.026$, $t = 2.724$, $p < .01$). There were no differences in RTs between SO-*Which* and OS-*Which*, neither in adults ($\beta = 0.025$, $SE = 0.035$, $t = 0.716$, $p = .474$) nor in children ($\beta = 0.006$, $SE = 0.033$, $t = 0.185$, $p = .853$).

Segment 4

The analysis of RTs at Segment 4 showed a main effect of Group ($\beta_{\text{Adults vs. Children}} = 0.229$, $SE = 0.085$, $t = 2.693$, $p = .009$) and a significant interaction between WhOrder and WhType ($\beta_{\text{SO vs. OS} \times \text{Who vs. Which}} = -0.073$, $SE = 0.036$, $t = -2.012$, $p = .044$). Pairwise comparisons indicate a significant difference between the SO and OS orders in *who*-questions, with SO shorter than OS ($\beta = -0.047$, $SE = 0.023$, $t = -2.011$, $p = .044$), but not in *which*-questions ($\beta = 0.026$, $SE = 0.028$, $t = 0.929$, $p = .353$). Further pairwise comparisons reveal that this difference in Segment 4 is driven by the fact that in adults, but not in children, SO-*Who* questions yielded significantly shorter RTs than OS-*Who* questions ($\beta = -0.077$, $SE = 0.032$, $t = -2.416$, $p = .015$).

Segment 5

We found again a significant main effect of Group, which reflects longer RTs in children as compared to adults ($\beta_{\text{Adults vs. Children}} = 0.226$, $SE = 0.079$, $t = 2.868$, $p < .01$).

Let us now compare the RTs for the OS conditions only in order to investigate whether the type of wh-element affects children's and adults' sensitivity to ordering constraints in multiple wh-questions. As above, we only report the significant effects and interactions and we present the data segment-by-segment. Figures 6 and 7 show the RTs for the OS conditions for children and adults, respectively.

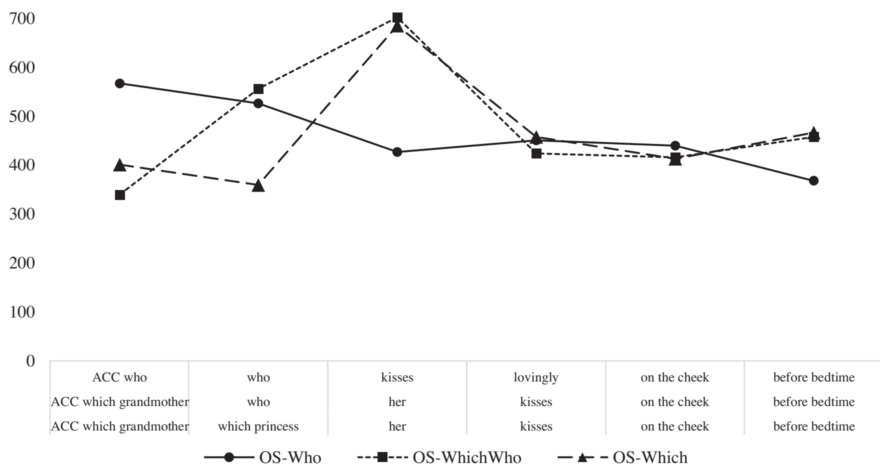

Figure 6. Distribution of children's RTs (in ms) for the different sentence segments (English translation) for multiple wh-questions with an objet-subject order (OS-*Who*, OS-*Which*, OS-*WhichWho*).

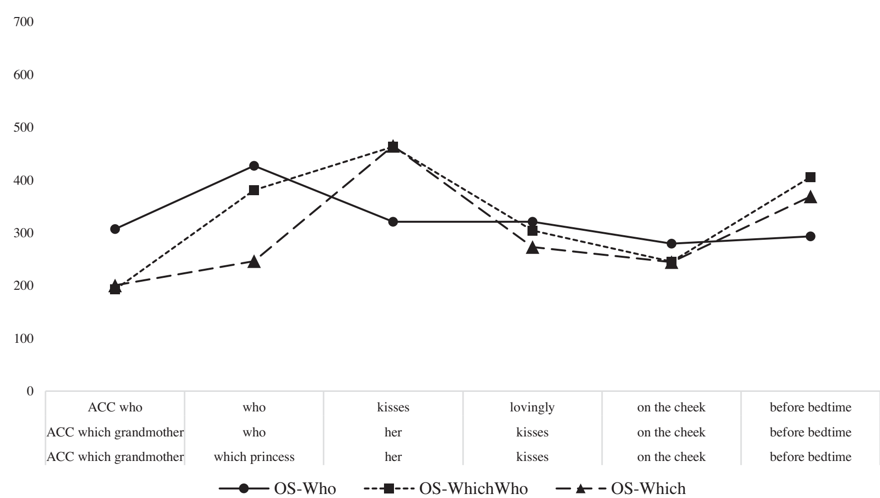

Figure 7. Distribution of adults' RTs (in ms) for the different sentence segments (English translation) for multiple wh-questions with an objet-subject order (OS-Who, OS-Which, OS-WhichWho).

Segment 1

The analysis for the OS conditions revealed a significant effect of Group at Segment 1 ($\beta_{\text{Adults vs. Children}} = 0.281$, $SE = 0.085$, $t = 3.281$, $p < .01$), which indicates longer RTs in children as compared to adults. We also found significant differences between the levels of the WhType factor. To recall, because WhType was a three-factor level in the case of the OS conditions, we specified the following contrasts in the model (c2 vs. 1, which tested the differences between the level *Who* and the level *WhichWho*, and c3 vs. 2, which tested the differences between the level *WhichWho* and the level *Which*). Therefore, the analysis reflects significantly shorter RTs for the *WhichWho* than the *Who* conditions at Segment 1 ($\beta_{\text{Who vs. WhichWho}} = -0.325$, $SE = 0.054$, $t = -5.975$, $p < .001$). The interaction between $\text{WhType}_{\text{Who vs. WhichWho}}$ and Group was also significant ($\beta_{\text{Who vs. WhichWho} \times \text{Adults vs. Children}} = -0.121$, $SE = 0.053$, $t = -2.270$, $p = .023$). Subsequent pairwise comparisons indicate shorter RTs for OS-*WhichWho* than for OS-*Who* in adults ($\beta_{\text{Who vs. WhichWho}} = -0.276$, $SE = 0.082$, $t = -3.344$, $p < .01$) and a more pronounced difference in children ($\beta_{\text{Who vs. WhichWho}} = -1.384$, $SE = 0.059$, $t = -6.470$, $p < .001$).

Segment 2

We found a significant effect of WhType ($\beta_{\text{WhichWho vs. Which}} = -0.386$, $SE = 4.773$, $t = -8.091$, $p < .001$), associated with significantly shorter RTs for *Which* than for *WhichWho* at Segment 2. Furthermore, there were two significant interactions between (a) $\text{WhType}_{\text{Who vs. WhichWho}}$ and Group ($\beta_{\text{Who vs. WhichWho} \times \text{Adults vs. Children}} = 1.891$, $SE = 5.256$, $t = 3.599$, $p < .001$) and (b) $\text{WhType}_{\text{WhichWho vs. Which}}$ and Group ($\beta_{\text{WhichWho vs. Which} \times \text{Adults vs. Children}} = -1.546$, $SE = 5.841$, $t = -2.648$, $p < .01$). These illustrate (a) shorter RTs for OS-*WhichWho* than for OS-*Who* only in adults ($\beta_{\text{Who vs. WhichWho}} = -0.122$, $SE = 0.055$, $t = -2.196$, $p = .041$) and (b) shorter RTs for OS-*Which* than for OS-*WhichWho* both in adults ($\beta_{\text{WhichWho vs. Which}} = -0.410$, $SE = 0.059$, $t = -6.933$, $p < .001$) and in children ($\beta_{\text{WhichWho vs. Which}} = -0.459$, $SE = 0.061$, $t = -7.422$, $p < .001$).

Segment 3

At Segment 3, there was a significant effect of Group ($\beta_{\text{Adults vs. Children}} = 2.009$, $SE = 7.334$, $t = 2.740$, $p < .01$) and a significant effect of WhType ($\beta_{\text{Who vs. WhichWho}} = 3.493$, $SE = 5.627$, $t = 6.207$, $p < .001$), which reflects significantly longer RTs for *WhichWho* than for *Who* at Segment 3. The clitic region thus yields longer RTs in the *WhichWho* trials as compared to the verb region in the *Who* trials. There was also a significant interaction between $\text{WhType}_{\text{Who vs. WhichWho}}$ and Group ($\beta_{\text{Who vs. WhichWho} \times \text{Adults vs. Children}} = 1.766$, $SE = 4.479$, $t = 3.943$, $p < .001$). Subsequent pairwise comparisons show longer RTs for OS-*WhichWho* than for OS-*Who* in adults ($\beta_{\text{Who vs. WhichWho}} = 0.378$, $SE = 0.081$, $t = 4.641$, $p < .001$) and in children ($\beta_{\text{Who vs. WhichWho}} = 0.437$, $SE = 0.069$, $t = 6.308$, $p < .001$).

Segment 4

The analysis revealed a main effect of Group ($\beta_{\text{Adults vs. Children}} = 0.166$, $SE = 0.077$, $t = 2.151$, $p = .036$).

Segment 5

The analysis revealed a main effect of Group ($\beta_{\text{Adults vs. Children}} = 2.504$, $SE = 7.807$, $t = 3.207$, $p < .01$).

Segment 6

The difference between the *Who* and *WhichWho* conditions only approaches significance ($\beta_{\text{Who vs. WhichWho}} = 0.150$, $SE = 0.073$, $t = 2.039$, $p = .051$) but goes in the same direction as in the preceding segments to show that the *WhichWho* condition yields longer RTs than the *Who* condition.

We also examined whether RTs at each segment vary as a function of Age in the Child group and whether Age modulates the online sensitivity to ordering constraints; however, no significant effect of Age emerged at any segment. In addition, we performed an analysis of RTs for the trials with incorrect responses in the SO and OS-*Which* conditions. The small amount of inaccurate trials for the other *Who* conditions and for the adult data (see Table 1) did not allow any analyses to be conducted for these conditions and for the adults. Visual inspection of the data, followed up by LMER models at each segment, for the *Which* conditions only, revealed no significant effect of wh-order. The OS-*Which* condition yielded faster RTs than the SO-*Which* condition at the last segment; however, the effect did not reach significance ($p = .142$).

Summarizing, the online reaction data reveal a slowdown for *who*- versus *which*-phrases, as well as longer RTs associated with the clitic region in multiple *which*-questions in both children and adults. Children also show longer RTs for wh-objects as compared to wh-subjects (irrespective of the type of wh-element). Adults show a slowdown in RTs in the OS-*Who* as compared to the SO-*Who* conditions.

Discussion

In this study, we aimed to examine the acquisition and processing of multiple *who* and *which*-questions in Romanian that display ordering constraints and involve exhaustivity. The specific goals were to determine (a) how children and adults process multiple *who*-questions as compared to *which*-questions, (b) whether they display an online sensitivity to ordering constraints in multiple wh-questions, and (c) the extent to which the type of wh-element affects the comprehension of multiple wh-questions.

We carried out a self-paced listening experiment that simultaneously investigated online processing and offline comprehension of multiple wh-questions in Romanian children and adults. The study manipulated the type of wh-phrase (*who* vs. *which*) and the order of these elements (SO vs. OS). Romanian requires all wh-phrases to be fronted and exhibits strict ordering constraints in *who*, but not in *which*-questions: fronting a *who*-object over a *who*-subject is ungrammatical, while fronting a *which*-object over a *which*-subject or a *who*-subject is not. Accuracy analyses tested the offline comprehension of multiple wh-questions and allowed to address (a) the impact of the type of wh-element on response accuracy, as this has been shown to affect the offline comprehension of single wh-questions and (b) the issue of exhaustivity because we measured whether children and adults used an exhaustive or a non-exhaustive response, in other words, whether they give the exhaustive sets of possible answers to both wh-elements, which are pairwise linked. RT analyses measured how children and adults process multiple *who*- and multiple *which*-questions online and whether they are sensitive to the ordering constraints present in multiple *who*, but not in multiple *which*-questions.

The findings for offline accuracy show that the type of wh-element (*who* vs. *which*) modulates the comprehension of multiple wh-questions. Both children and adults comprehend *who*-questions very well, even when the object fronts over the subject, despite the fact that both the object and the subject in the OS condition share the same +Q or interrogative feature. To recall, an analysis of multiple wh-questions in terms of featural intervention and RM postulates that multiple *who*-questions with an OS order involve a featural identity relation that should be ruled out as ungrammatical because the elements entering this featural relation appear in a nesting chain configuration (Krapova & Cinque, 2008; Rizzi, 2011, 2017). However, the findings reveal that children and adults have no difficulties comprehending the OS-*Who* ungrammatical sentences. We attribute this result to the specificity of the task: participants are required to listen to the question and then map its meaning unto a correct interpretation in order to select all the pictures that match the described action. In so doing, the participants need to encode back-to-back wh-elements and then, upon reaching the verb, retrieve them and establish the correct thematic relations. Both children and adults are able to map the wh-phrases onto the corresponding thematic relations, while disregarding the ungrammaticality of the sentence. In other words, they can successfully repair the ungrammatical structure in order to arrive at the correct interpretation for the question. Note that these results are in line with those of Hofmeister et al. (2013), as these authors also report high accuracy scores for multiple *who*-questions with an ungrammatical order (84.4% accuracy), although in their study questions with two bare wh-words are less accurate than questions with two *which*-phrases.

Over-exhaustive responses constituted the only type of error that the adults in our study made when answering multiple *who*-questions (Table 4). Thus, on rare occasions, they also identify the pair of referents performing the wrong action. Nonetheless, the fact that they do not give any single pair answers or exhaustive lists of subjects or objects is evidence that they do not have difficulties with any of the two steps necessary to derive a paired list answer: exhausting the question domain and pairing the two wh-elements. Children also give over-exhaustive answers to multiple *who*-questions, but mostly they make singleton response errors (Table 4), in which they provide exhaustive lists of referents either for the wh-subject or for the wh-object. A closer inspection of error types for each age group (Table 5) reveals that this is the most frequent type of error in the youngest children tested, the 6- to 7-year-olds. This is in line with previous studies on the comprehension of multiple wh-questions in Romanian (Bentea, 2010; Măniță, 2017) showing that even at the age of 6, Romanian children tend to answer only one of the wh-words in the question. However, when answering only one wh-word, the children in our study give exhaustive lists of subjects and objects which can indicate that they only exhaust over one wh-phrase, while the other might not be present in their interpretation (see Schulz & Roeper, 2011). Our results therefore reveal that, at the age of 8, children interpret multiple wh-questions as requiring exhaustive PL answers. Younger children, on the other hand, seem to have acquired the exhaustive reading, but they have difficulties linking the two wh-words.

In the case of multiple *which*-questions, children and adults are more accurate with questions involving a SO order as compared to those in which the object moves across the subject. For children, we found an effect of age showing that younger children give less accurate responses than older children to SO-*Which* questions. Children and adults have more difficulties assigning a correct interpretation to OS-*Which* questions. On the other hand, children differ from adults in their comprehension of OS-*WhichWho* questions, that is, questions with an OS order and in which the object contains a lexical N restriction, while the subject is a bare wh-word (*who*). Whereas adults comprehended OS-*WhichWho* very well, on a par with OS-*Who* questions, children struggle more with the comprehension of OS-*WhichWho* questions. At first view, this result follows from an account of multiple wh-questions in terms of featural intervention, given that both OS-*Which* (16) and OS-*WhichWho* (17) questions instantiate an inclusion relation created by the presence of a +Q feature on the two fronted wh-elements:

+Q +N +Q

16. Pe care bunică care fată <care fată> o sărută <pe care bunică>?

'Which grandmother is which girl kissing?'

+Q +N +Q

17. Pe care bunică cine <cine> o sărută <pe care bunică>?

'Which grandmother is who kissing?'

This could indicate that children find it difficult to comprehend or compute inclusion configurations in general and not just those created by a +N feature (or lexical restriction) shared between the moved element and the intervening one, as in the case of single *which*-object questions. Moreover, given an analysis which postulates two potential attractors (+Q +N and +Q) for lexically restricted wh-phrases (Villata et al., 2016, Rizzi, 2017), (17) is predicted to yield similar results to (16) because both give rise to the same inclusion configuration in which the +Q feature on the two interveners (the nested chains formed by the moved subjects *care fată* [which girl] and *cine* [who] and their traces) is included in the featural specification of the element that gets attracted to a higher +Q +N position.

The results show, however, that both children and adults comprehend OS-*WhichWho* questions (17) significantly better than OS-*Which* questions (16). Therefore, the two cases of inclusion cannot be considered on a par. We postulate that the featural similarity in lexical N restriction between the two wh-arguments in OS-*Which* questions drives the added complexity in comprehension. The presence of a wh-subject containing a lexical N restriction in OS-*Which* questions hinders the correct assignment of thematic relations to the two wh-elements which, in turn, leads to misinterpret OS-*Which* questions as containing a SO order instead. The finding that even SO-*Which* questions pose more difficulties for younger children than for older children also suggests that a lexical N restriction on both the wh-subject and the wh-object makes *which*-questions harder for comprehension, despite the presence of case information. One possibility is that children's overall difficulties with multiple *which*-questions could be related to the presence of two lexically restricted wh-words. This could make parsing the question in relation to the visual cues more taxing, given that the visual cues were presented at the end, after participants heard the whole sentence. However, if this were the case, then we would expect children to comprehend SO-*Which* and OS-*Which* questions on a par, as they both contain an additional lexically restricted element. The fact that we find an asymmetry in comprehension, with lower accuracy for structures in which the object precedes the subject, suggests that the difficulties associated with multiple *which*-questions containing an OS order stem from intervention effects triggered by the presence of a lexical +N restriction in the set of features characterizing both the wh-object and the intervening subject chain.

The difficulty with OS-*Which* questions is reflected in type of errors that participants make for multiple *which*-questions, as the most frequent errors are role reversals, meaning that both children and adults interpret the first NP as Agent and the second NP as Patient. Although *which*-questions are harder for children to comprehend than *who*-questions, this does not show that they have not acquired exhaustivity in *which*-questions. The other errors children make include over-exhaustive responses (i.e., children answer by listing all the pairs of characters in the visual display, including the one in which the Agent–Patient roles are reversed) and singleton answers (again, children provide exhaustive lists of subjects or objects). We take these errors to suggest that children show mastery of exhaustivity in multiple *which*-questions as well. In addition, the most common errors in *which*-questions were reversals, which we do not consider on a par with single-pair answers, which would have been equivalent to answering with only one of the pairs performing the correct action. Rather, children's errors with *which*-questions reveal that children have difficulties assigning the correct theta-roles when both wh-arguments contain a nominal restriction.

Intervention effects in comprehension arise despite the fact that *which*-objects in Romanian are marked for Case by the preposition *pe*. This is in line with the literature that has tested the effect of Case mismatch on the comprehension of single object *which*-questions (for Hebrew: Friedmann et al., 2017; for Romanian: Bentea, 2016) showing that the dissimilarity in Case features between the moved object and the intervening subject does not enhance the comprehension of object *which*-questions and cannot overcome the intervention effects found in these structures. Friedmann et al. (2017), following Belletti et al. (2012), argue that only mismatches in features acting as triggers of syntactic movement, typically inflected on the verb, can facilitate intervention configurations and that Case, although relevant for movement, does not trigger it, and thus is not a relevant feature for modulating intervention effects.[11]

Hence, the inclusion of a +N feature seems to be more penalizing for comprehension than the inclusion of a +Q feature alone. Our data thus suggest that an analysis in which the second *which*-phrase is attracted by a simple +Q head does not fully account for the response pattern obtained for OS-*Which* and OS-*WhichWho* questions. However, if a +N feature is also present on the lower *which*-element, we are now faced with the challenge of accounting for the grammaticality of these structures, since now both elements are specified for the same features (+Q +N) and the configuration yields nested chains. A syntactic analysis of multiple wh-movement goes beyond the scope of this paper, but we could speculate that one possibility to derive the grammaticality of examples like (16) would be to broaden the class of features to also include a +Top(ic) feature on the *which*-phrases, Top being the feature associated with D(iscourse)-linking. This proposal runs into issues of its own. Under the assumption that D-linking is determined in the presence of a context of utterance, then even the *Who*-questions in our study are D-linked and thus specified for a +Top feature since all the questions were preceded by a lead-in introducing a specified set of referents. Another possibility would be to include both +Top and a different featural specification (potentially captured in terms of *specificity*) for clitic-resumed and non-clitic-resumed Topics. This would go in line with Krapova and Cinque (2008, p. 186) who show that, at least in Bulgarian, clitic-resumed *which*-phrases target a different position than non-clitic-resumed *which*-phrases.

Another potential explanation for the results obtained comes from cue-based interference models (Lewis & Vasishth, 2005; Lewis et al., 2006; van Dyke & McElree, 2011), which account for difficulties with long-distance dependency processing in terms of constraints from memory retrieval mechanisms. Under this view, memory retrieval is driven by cues, which identify the features of the element(s) to be retrieved and distinguish it from other irrelevant representations in memory. Specifically, upon encountering a constituent (e.g., the wh-phrase in a question), information about this element is encoded in memory (e.g., syntactic category, animacy, argument, etc.). This constituent then has to be retrieved from memory at the gap position and integrated into the structure. At this point, the previously encoded cues are analyzed and if another constituent shares similar cues with those of the element that needs to be retrieved from memory, this second constituent will interfere with the processing of the initially encoded element. This then results in an increased processing cost for the structure. When the cues of potentially interfering constituents are sufficiently different, this results in a reduced processing cost for the structure, which, in turn, will make the structure easier to comprehend. In the multiple wh-questions tested in this study, not just one but two wh-constituents need to be encoded at the very beginning of the structure and then the information encoded needs to be maintained in memory until it can be retrieved at the gap positions where these constituents can be successfully integrated into the structure. If the set of cues of the subject and object in a multiple *which*-question are sufficiently similar, this will overload memory capacity and the structure will be more costly for comprehension. If the two sets of cues are dissimilar, like in an OS-*WhichWho* question, memory resources will be less burdened and the structure easier to comprehend. Further research is needed to assess whether children's memory skills interact with their language abilities to modulate the comprehension of complex structures like multiple *which*-questions. This would require the use of a working memory task, as well as a comparison of different experimental designs, not only designs where the pictures appear after the end of the sentences, like in the present study, but also designs where the pictures are present on screen while the sentences are being processed.

Moving on to the analyses of RTs, these reveal an effect of type of wh-element in both groups of participants, with shorter RTs when processing *which*- as compared to *who*-elements. This effect surfaces when participants encode the syntactic and semantic information associated with the wh-fillers, so before they reach the verb region where they have to retrieve this information and successfully map the wh-phrases to the thematic structure of the verb. This is in line with the self-paced reading results reported in Hofmeister et al. (2013). Although not directly comparable, as Hofmeister et al. (2013) do not report word-by-word reading times for the whole sentence, their results reveal shorter residual reading times at the word immediately preceding the verb when this is a *which*-phrase compared to when it is a *who*-phrase. However, contrary to the predictions based on Hofmeister et al. (2013), we did not find a difference in RTs at the verb (nor its spillover regions) between the conditions with two *who* and those with two *which* elements, as participants do not process the verb region faster in the conditions with one or two *which*-phrases.

Moreover, the online results reveal that adults, but not children, listen longer to the wh-subject in the OS-*Who* conditions, so those conditions which violate Superiority constraints. In the SO conditions, participants heard a wh-subject in Segment 1 followed by a wh-object in Segment 2, whereas in the OS conditions they heard a wh-object in Segment 1, followed by a wh-subject in Segment 2. While a *which*-object preceding a *which*-subject or a *who*-subject is grammatical, a *who*-object preceding a *who*-subject leads to Superiority violations and the sentence should be ruled out as ungrammatical. Adults, but not children, show an online sensitivity to ordering constraints in multiple *who*-questions. The advantage of a processing system that is fully developed in adults could account for the fact that adults, but not children, can detect the ungrammaticality of OS-*Who* questions. Children, unlike adults, are unable to recruit this information in real-time processing or it could be that the timing of the effect might take much longer to surface in children. There is evidence that 5-year-olds do not actively form filler-gap dependencies in real-time comprehension of wh-questions and that when active dependency formation appears in 6-year-olds, there is a small delay in its execution as compared to adults (Atkinson et al., 2018). Other visual world studies have also found young children to be slow in processing filler-gap dependencies, with effects occurring after the end of the sentence (Adani & Fritzsche, 2015). The children in our study are older; however, the structures tested are more complex as they require encoding and integrating two wh-elements in the structure. Another possibility is that the ordering constraint in children seems to be overridden by the quest for meaning. It could be a task effect as children first have to listen and get the correct meaning of the questions they are hearing, and then they have to select the correct pictures. Children can do that by repairing the ungrammaticality of the sentence, given that there is sufficient time between the moment when the wh-dependency is initiated and the moment when it is interpreted. When children are given time to encode the meaning of each scene before giving an answer, they can plausibly understand the sentence and map the wh-phrases unto the correct argument structure of the verb. Further research including finer-grained measures of sentence processing, like visual world paradigm, as well as a production task, could provide additional evidence that children's online insensitivity to ordering constraints is due to a task effect or related to a non-adult-like grammar.[12]

We also found longer RTs associated with the clitic region in multiple *which*-questions in both children and adults, which indicates that participants have more difficulties processing these elements. Clitic doubling requires extra processing because upon encountering the clitic, one needs to identify the correct antecedent, namely the wh-object. The clitic appears in a derived position preceding the verb (see Coene & Avram, 2012 for analyses of clitic doubling constructions in

Romanian) and requires the establishment of an additional syntactic and referential dependency with its antecedent. Children might find this more difficult than adults, although both groups slow down when processing the clitic, which, in turn, results in lower accuracy scores for multiple questions containing a *which*-object. Moreover, the fact that the two potential antecedents, the wh-subject and the wh-object, match in gender and number features could render the processing of the clitic more costly for children and adults alike.

To conclude, our findings indicate a speed-accuracy trade-off. Children and adults are more accurate with multiple *who*- than *which*-questions, but they slow down when they process *who*- as compared to *which*-phrases. An intervention effect appears in OS-*Which* questions but only in accuracy, showing that participants find it harder to establish the correct thematic relations between the moved wh-phrases and the verb in the presence of two *which*-phrases. We identified the source of this intervention effect as the inclusion of a lexical +N restriction in the set of features characterizing both the wh-object and the intervening subject chain. This inclusion of a +N feature seems to be more costly for comprehension than the inclusion of a +Q feature alone, because children and adults comprehend OS-*WhichWho* questions better than *OS-Which* questions. The lack of intervention effects in terms of RTs indicates that such effects occur at a later stage, after children have heard the whole sentence and when they interpret its meaning.

Acknowledgements. The research presented here was supported by a grant from the Swiss National Science Foundation (project P2GEP1_174870) awarded to AB. We thank all the participating children and their parents, all the adult participants, Peter Hudspith for drawing the pictures, as well as the editor and three anonymous reviewers for insightful comments and helpful suggestions.

Notes

1. The [+Q] feature designates the presence of a question operator.

2. Subsequent renditions of this constraint, for example, Attract Closest or the Minimal Link Condition (Chomsky, 1995, 2000), retain the same gist.

3. Pesetsky (1987) introduces the notion of "Discourse-linked" (D-linked) to refer to wh-phrases like *which cat* or *which book*, as these elements prompt an answer chosen from a set of referents already present in the discourse, whereas wh-phrases like *who* do not.

1. "*Which book* did *who* buy?"

4. As a reviewer points out, it is not clear why some languages such as German do not exhibit Superiority effects, or at least not as strong as in English, with bare wh-words either (*Was kauft wer?* "What bought who?") if the lack of Superiority effects can be accounted for in terms of D-linking. Indeed, Pesetsky (2000) uses the derivation put forth for English *which* or D-linked questions to explain the structure of all German questions and therefore postulates that "in a German multiple wh-question, all wh-in-situ undergo wh-feature movement" (2000:72) (see Kotek, 2019 for an alternative account). Grohmann (2002) also proposes that apparent superiority violations in German involve D-linking. However, these accounts cannot predict findings from processing of multiple wh-questions in German (Featherston, 2005) which indicate an ordering preference consistent with the one reported for English: Superiority violating sentences like (1) were judged less acceptable than structures obeying Superiority like (2):

1. Maria fragt wer was gelesen hat.
   Maria asks who what read has
   "Maria asks who has read what."
2. Maria fragt was wer gelesen hat.
   Maria asks what who read has
   "Maria asks what who has read."

5 In addition, there is evidence from the literature on adult sentence processing that *which*-elements, compared to *who*-elements, increase the acceptability of sentences with island violations (for English: Goodall, 2014; Atkinson et al., 2016; for French: Villata et al., 2016).
6 On the other hand, in a self-paced reading study in Dutch with single *who*- and *which*-questions with role reversal (1), Donkers, Hoeks, & Stowe (2013) (see also references therein) found that, compared to *who*, the *which N* questions showed consistently longer reading times until the final segment of the sentence.

1 *Who/Which servant* has the emperor looked for in the cellar?

7. Note that multiple wh-questions containing two *which*-elements in English (1) can be answered both with (1a), a pair-list answer, and with (1b), a single-pair answer:

1. Which child bought which book?
(a) John bought *Zog* and Mary bought *The Gruffalo*.
(b) John bought *Zog*.

However, the availability of single-pair answers to questions like (1a) remains an open issue. Some authors (Barss, 2000; Dayal, 2016) find them acceptable, whereas others (Comorovski, 1996) consider multiple *which*-questions unacceptable on the single-pair reading.
8. *Which*-elements can also be separated from other bare wh-phrases by fronting them in a matrix clause with bare elements appearing in a lower position (1). This option is ruled out for bare wh-words (2).

1.	*Pe care student*$_i$	vrei	să	ştii	*cine*	*când*	l$_i$-a	intervievat?
	ACC.which student$_i$	want.2.SG	SUBJ	know.2.SG	who	when	him$_i$-has	interviewed
	"Which student do you want to know who interviewed when?"							
2.	**Pe cine*	vrei	să	ştii	*cine*	*când*	a	intervievat?
	ACC.who	want.2.SG	SUBJ	know.2.SG	who	when	has	interviewed
	"Whom do you want to know who interviewed when?"							

9. One reviewer notes that the study did not include sentences with one bare phrase and one *which*-phrase that violate ordering constraints. Sentences with two *who*-phrases represent a clear case of ungrammaticality and are also consistently judged as highly degraded in acceptability judgment studies with superiority violations (Hofmeister et al., 2013) and extractions from weak islands (Villata et al., 2016). Sentences where *who* precedes *which*, on the other hand, are judged to be significantly better than those with two *who*-elements and significantly worse than those with two *which*-elements (Hofmeister et al., 2013). Informal judgments from adult Romanian-speakers seem to confirm this pattern for Romanian as well. Given the gradient in judgments associated with ungrammatical sentences in which *who* precedes *which*, we have decided not to include them in the present study. However, this paves the way for a follow-up study assessing sensitivity to ordering constraints in questions that contain both a *who*-phrase and a *which*-phrase either in subject or in object position.
10. We only used 10 fillers in order to reduce the length of the task itself and of the test sessions. We could only take the children out of their classroom for 45 min at a time and especially younger children found it difficult to concentrate for more than 30 min.
11. However, divergent findings are reported in other studies (see, e.g., Varlokosta, Nerantzini & Papadopoulou (2015), who looked at the comprehension of movement structures like wh-questions and relative clauses in Greek-speaking children).
12. One anonymous reviewer suggests that the result showing adults, but not children, to be sensitive to the ordering violation in OS-*Who* questions, could be more a question of metalinguistic awareness rather than a focus on meaning. Although this is a plausible interpretation, the current design of the experiment does not allow us to directly address this. This remains for future research. One possibility would be to use a grammaticality judgment task, along the lines of Gavarró (2020), who examined children and adults' judgments of object relative clauses, long-distance wh-questions, and ungrammatical wh-questions involving RM

violations in Catalan and found that children reject sentences containing such violations more often than object relatives or long-distance wh-questions.

13. The total number of answers in the *who*-conditions is 319 as there is one missing value both for SO-*Who* and for OS-*Who*.

References

Adani, F., & Fritzsche, T. (2015). On the relation between implicit and explicit measures of child language development: evidence from relative clause processing in 4-year-olds and adults, in E. Grillo & K. Jepson (Eds.) *Proceedings of the 39th Annual Boston University Conference on Language Development* (pp. 14–26). Somerville, MA: Cascadilla Press.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control,* **19**, 716–723.

Alboiu, G. (2000). *The Features of Movement in Romanian.* Bucharest: EUB.

Avrutin, S. (2000). Comprehension of Wh-questions by children and Broca's aphasics. In Y. Grodzinsky, E. P. Shapiro, & D. A. Swinney (eds.), *Language and the brain: Representation and processing* (pp. 295–312). San Diego, CA: Academic Press.

Atkinson, E., Apple, A., Rawlins, K., & Omaki, A. (2016). Similarity of wh-phrases and acceptability variation in wh-islands. *Frontiers in Psychology*, **6**, 1–16. https://doi.org/10.3389/fpsyg.2015.02048

Atkinson, E., Wagers, M. W., Lidz, J., Phillips, C., & Omaki, A. (2018). Developing incrementality in filler-gap dependency processing. *Cognition*, **179**, 132–149.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, **59**(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Barss, A. (2000). Minimalism and asymmetric wh-interpretation. In R. Martin, D. Michaels, J. Uriagereka, & S. J. Kaiser (Eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik* (pp. 31–52). Cambridge, MA: MIT Press.

Bates, D. M., Kliegl, R., Vasishth, S., & Baayen, R. H. (2018). Parsimonious Mixed Models. arXiv:1506.04967

Bates, D. M., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, **67**(1). https://doi.org/10.18637/jss.v067.i01

Belletti, A., Friedmann, N., Brunato, D., & Rizzi, L. (2012). Does gender make a difference? Comparing the effect of gender on children's comprehension of relative clauses in Hebrew and Italian. *Lingua*, **122**(10), 1053–1069. https://doi.org/10.1016/j.lingua.2012.02.007

Bentea, A. (2010). Multiple Wh-questions in Child Language:A Cross-Linguistic Perspective. In P. Guijarro-Fuentes & L. Dominguez (Eds.), *New Directions in Language Acquisition: Romance Languages in the Generative Perspective* (pp. 1–485). Newcastle upon Tyne: Cambridge Scholars Publishing.

Bentea, A. & Durrleman, S. (2013) A'-dependencies in French: A Study in L1 Acquisition. In S. Baauw, F. A.C. Drijkoningen, L. Meroni, and M. Pinto (Eds). *Romance Languages and Linguistic Theory 2011. Selected papers from 'Going Romance' Utrecht 2011* (pp. 1–16). Amsterdam: John Benjamins.

Bentea, A. (2016). Comprehension of wh-questions in child Romanian : A case about case and lexical restriction. In J. Choi, H. Demirdash, O. Lungu, & L. Voeltzel (Eds.), *Language Acquisition at the Interfaces: Proceedings of GALA 2015* (pp. 1–18). Newcastle upon Tyne: Cambridge Scholars Publishing.

Boškovic, ž. (2002). On multiple wh-fronting. *Linguistic Inquiry 33.* 351–384.

Chomsky, N. (1973). Conditions on transformations. In S. Anderson & P. Kiparsky (Eds.), *A Festschrift for Morris Halle* (pp. 232–286). New York: Holt, Reinhart & Winston.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Chomsky, N. (2000). Minimalist inquiries: the framework. In R. Martin, D. Michaels, J. Uriagereka, & S. J. Kaiser (Eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik* (88–155). Cambridge, MA: MIT Press.

Chondrogianni, V., Marinis, T., Edwards, S., & Blom, E. (2014). Production and on-line comprehension of definite articles and clitic pronouns by Greek sequential bilingual children and monolingual children with specific language impairment. *Applied Psycholinguistics*, **36**(5), 1155–1191. https://doi.org/10.1017/S0142716414000101

Coene, M., & Avram, L. (2012) "(A)symmetries in the acquisition of Romanian clitics". in: P. Guijarro-Fuentes & M. P. Larranaga (Eds.), *Pronouns and clitics in early language,* pp 131–173, Berlin: Mouton de Gruyter.

Comorovski, I. (1989). Discourse and the syntax of multiple constituent questions (PhD thesis). Cornell University, Ithaca, NY.

Comorovski, I. (1996). *Interrogative phrases and the syntax-semantics interface*. Norwell, MA: Kluwer Academic Publishers.

Contemori, C., Carlson, M., & Marinis, T. (2018). On-line processing of English which -questions by children and adults: a visual world paradigm study. *Journal of Child Language*, **45**(2), 415–441. https://doi.org/10.1017/s0305000917000277

Contemori, C., & Marinis, T. (2014). The impact of number mismatch and passives on the real-time processing of relative clauses. *Journal of Child Language*, **41**(3), 658–689. https://doi.org/10.1017/S0305000913000172

Dayal, V. (2002). Single-pair versus multiple-pair answers: Wh-in-situ and scope. *Linguistic Inquiry*, **33**(3), 512–520.

Dayal, V. (2016). *Questions*. New York: Oxford University Press.

De Vincenzi, M. (1991). *Syntactic parsing strategies in Italian: The minimal chain principle*. Dordrecht, The Netherlands: Kluwer.

De Vincenzi, M., Arduino, L. S., Cicarelli, L., & Job, R. (1999). Parsing strategies in children comprehension of interrogative sentences. *Proceedings of European Conference in Cognitive Science* (pp. 301–308). Siena: Università di Siena.

Dobrovie-Sorin, C. (1994) *The Syntax of Romanian: comparative studies in Romance*. Berlin: Mouton de Gruyter.

Donkers, J., Hoeks, J. C. J., & Stowe, L. A. (2013). D-linking or set-restriction? Processing Which-questions in Dutch. *Language and Cognitive Processes*, **28**(1–2), 9–28. https://doi.org/10.1080/01690965.2011.566343

Featherston S. (2005). Universals and grammaticality: Wh-constraints in German and English. *Linguistics* **43**:4.

Felser, C., Marinis, T., & Clahsen, H. (2003). Children's Processing of Ambiguous Sentences: A Study of Relative Clause Attachment. *Language Acquisition*, **11**(3), 127–163. https://doi.org/10.1207/s15327817la1105_1

Friedmann, N., Belletti, A., & Rizzi, L. (2009). Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua*, **119**(1), 67–88. https://doi.org/10.1016/j.lingua.2008.09.002

Friedmann, N., Rizzi, L., & Belletti, A. (2017). No case for Case in locality: Case does not help interpretation when intervention blocks A-bar chains. *Glossa: A Journal of General Linguistics*, **2**(1), 33. https://doi.org/10.5334/gjgl.165

Gavarró, A., Lewandowski, W., & Markova, A. (2010). 'An Approach to Multiple Interrogatives in Child Bulgarian and Polish'. In J. Costa, A. Castro, M. Lobo, & F. Prata (Eds.) *Language Acquisition and Development Proceedings of GALA 2009*, 170–181, Newcastle upon Tyne: Cambridge Scholars Publishing.

Gavarró, A. (2020). Child Relativized Minimality and Grammaticality Judgement. *Frontiers in Psychology*, **11**:105. doi: 10.3389/fpsyg.2020.00105

Goodall, G. (2014). The D-linking effect on extraction from islands and non-islands. *Frontiers in Psychology*, **5**, 1–11. https://doi.org/10.3389/fpsyg.2014.01493

Goodluck, H. (2005). D(iscourse)-Linking and question formation: Comprehension effects in children and Broca's aphasics. In A. M. Di Sciullo (Ed.), *UG and External Systems* (pp. 185–192). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/la.75

Goodluck, H. (2010). Object extraction is not subject to Child Relativized Minimality. *Lingua*, **120**(6), 1516–1521. https://doi.org/10.1016/j.lingua.2009.10.005

Grebenyova, L. (2006). Multiple interrogatives in child Russian. In D. Bamman, T. Magnitskaia & C. Zaller (Eds.) *Proceedings of the 30th annual Boston University Conference on Language Development* (pp. 225–236). Sommerville: Cascadilla.

Grebenyova, L. (2011). Acquisition of multiple questions in English, Russian, and Malayalam. *Language Acquisition*, **18**(3), 139–175. https://doi.org/10.1080/10489223.2011.580672

Grohmann, K. (2002). Some concepts and consequences of discourse-restricted quantification. In M. Feist, S. Fix, J. Hay, & J. Moore (Eds.) *Proceedings of SCIL10. MIT Working Papers in Linguistics 37*, Cambridge, MA: MIT Press.

Grohmann, K. (2003). German is a multiple wh-fronting language. In C. Boeckx & K. Grohmann (Eds.), *Multiple wh-fronting* (pp. 99–130). Philadelphia: John Benjamins.

Hirsch, C. & Hartman, J. (2006). Some (wh-) questions concerning passive interactions. In A. Belletti, E. Bennati, C., Chesi, E., Di Domenico, E., and I., Ferrari (Eds.) *Proceedings of the Conference on Generative Approaches to Language Acquisition*. Cambridge, UK: Cambridge Scholars Press.

Hofmeister, P., Jaeger, T. F., Arnon, I., Sag, I. A., & Snider, N. (2013). The source ambiguity problem: Distinguishing the effects of grammar and processing on acceptability judgments. *Language and Cognitive Processes*, **28**(1–2), 48–87. https://doi.org/10.1080/01690965.2011.572401

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, **59**(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Karttunen, L. (1977). Syntax and semantics of questions. *Linguistics & Philosophy*, **1**, 3–44.

Kotek, H. (2019). Composing Questions. *Linguistics Inquiry Monograph series*. Cambridge, MA: MIT Press.

Krapova, I., & Cinque, G. (2008). On the Order of wh-Phrases in Bulgarian multiple wh- Fronting. In G. Zybatow (Ed.) *Formal description of Slavic languages: the fifth conference, Leipzig 2003* (pp. 318-336). Frankfurt am Main: Peter Lang.

Krifka, M., 2001. For a structured meaning account of questions and answers. In Fery, C. & Sternefeld, W. (Eds.), *Audiatur vox sapentia. A Festschrift for Arnim von Stechow* (pp 287–319). Berlin: Akademie Verlag.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, **82**(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Laenzlinger, C., & Soare, G. (2005). A Cartographic Approach to Wh-movement in Romanian. *Working Papers in Linguistics*, **1**, 23–60.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* **29**, 375–419. doi: 10.1207/s15516709cog0000_25

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science* **10**(10), 447–454. doi: 10.1016/j.tics.2006.08.007

Marinis, T. (2010). Using on-line processing methods in language acquisition research. In Elma Blom & Sharon Unsworth (Eds.), *Language learning & language teaching: Vol. 27. Experimental methods in language acquisition research* (pp. 139–162). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/lllt.27.09mar

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, R. H., & Bates, D. M. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, **94**, 305–315. https://doi.org/10.1016/J.JML.2017.01.001

Măniță, C. (2017), Asymmetries in the Acquisition of Wh-questions: The Role of Intervention and Interference. Doctoral dissertation, University of Bucharest, Bucharest.

Meng, M., & Bader, M. (2000). Mode of disambiguation and garden-path strength: An investigation of subject-object ambiguities in German. *Language and Speech*, **43**, 43–74.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, **51**(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Peirce, J., & MacAskill, M. (2018). *Building experiments in PsychoPy*. London: Sage Publications.

Pesetsky, D. (1987). Wh-in-situ: movement and unselective binding. In E. Reuland & A. ter Meulen (Eds.), *The representation of (in)definiteness* (pp. 98–129). Cambridge, MA: MIT Press.

Pesetsky, D. (2000). *Phrasal movement and its kin*. Cambridge, MA: MIT Press.

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria.

Richards, N. 1997. What moves where when in which language? Doctoral dissertation, MIT, Cambridge, MA. http://www.ai.mit.edu/projects/dm/theses/richards97.pdf

Rizzi, L. (1990). *Relativized minimality*. Cambridge, MA: MIT Press.

Rizzi, L. (2004). Locality and the left periphery. In A. Belletti (Ed.), *Structures and beyond: The cartography of syntactic structures* (Vol. **3**, pp. 223–251). New York: Oxford University Press.

Rizzi, L. (2011). Minimality. In C. Boeckx (Ed.), *The Oxford handbook of linguistic minimalism* (pp. 220–238). New York : Oxford University Press.

Rizzi, L. (2013). Locality. *Lingua*, **130**, 169–186.

Rizzi, L. (2017). Comparing extractions from wh-islands and superiority effects. *Wiener Linguistische Gazette (WLG)*, **82**, 253–261.

Rizzi, L. (2018). Intervention effects in grammar and language acquisition. *Probus*, **30**(2), 339–367. https://doi.org/10.1515/probus-2018-0006

Roeper, T. & de Villiers, J. (1991). The emergence of bound variable structures. In Thomas L. Maxfield & Bernadette Plunkett (eds.), *Papers in the acquisition of WH: Proceedings of the University of Massachusetts Roundtable (UMOP)*, 225–266. Amherst: University of Massachusetts, GLSA.

Roeper, T. & de Villiers, J. 2000. The Acquisition Path for Wh-Questions. In de Villiers, J. & Roeper, T. (Eds.), *Handbook to Generative Approaches to Language Acquisition* (pp. 189–247). Berlin: Springer.

Roeper, T., Schulz, P., Pearson, B. Z., & Reckling, I. (2007). From singleton to exhaustive: The acquisition of wh-. In Michael Becker & Andrew McKenzie (eds.), *Proceedings of SULA 3: Semantics of Under-Represented Languages in the Americas*, 87–102. Amherst: University of Massachusetts, GLSA.

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2019). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. arXiv:1807.10451v4

Schlesewsky, M., Fanselow, G., Kliegl, R., & Krems, J. (2000). The subject preference in the processing of locally ambiguous Wh-questions in German. In B. Hemforth & L. Konieszny (Eds.), *German sentence processing* (pp. 65–93). Dordrecht: Kluwer.

Schulz, P. (2010). Who answered what to whom? On children's understanding of exhaustive questions. In COSTA33, London.

Schulz, P. & Roeper, T. 2011. Acquisition of exhaustivity in wh-questions: A semantic dimension of SLI?, *Lingua* **121**, 383–407.

Soare, G. (2009). The Syntax-Information Structure Interface: A Comparative View from Romanian. Doctoral dissertation, University of Geneva, Geneva.

Starke, M. (2001). Move dissolves into Merge: A theory of locality. Doctoral dissertation, University of Geneva, Geneva.

Stowe, L.A. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language & Cognitive Processes*, **1**, 227–245.

Tyack, D., & Ingram, D. (1977). Children production and comprehension of wh-questions. *Journal of child language*, **4**, 211–224.

van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language.* **65**, 247–263. doi: 10.1016/j.jml.2011.05.002

Varlokosta, S., Nerantzini, M., & Papadopoulou, D. (2015). Comprehension asymmetries in language acquisition: A test for Relativized Minimality. *Journal of Child Language*, (42), 618–661. https://doi.org/10.1017/S0305000914000257

Villata, S., Rizzi, L., & Franck, J. (2016). Intervention effects and Relativized Minimality: New experimental evidence from graded judgments. *Lingua*, **179**, 76–96. https://doi.org/10.1016/J.LINGUA.2016.03.004

Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. Berlin: Springer. https://doi.org/10.1007/978-3-319-24277-4

Appendix A

Task instructions and English translation (in parentheses we indicate what images appeared on the screen together with each set of instructions).

Acesta este un joc despre prințese și supereroi. Vei vedea imagini cu Elsa, Anna, Jasmine, Batman, Superman și Spiderman. Lor le place să se joace cu multe animale și cu mulți oameni. (*images of Elsa, Anna, Jasmine, Batman, Superman and Spiderman; whenever each name was mentioned, the experimenter pointed to the corresponding character to ensure the participants get familiarised with the names of the characters*)

Ursul Paddington îți va pune o întrebare. (*image of Paddington the Bear*)

Va trebui să apeși această tastă ca să auzi întrebarea lui Paddington. (*image of the "space" key*)

Te rog să asculți întrebarea cu foarte multă atenție. (*picture of ear and question mark*)

Vei vedea apoi pe ecran o imagine ca aceasta și va trebui să răspunzi la întrebarea lui Paddington bazându-te pe imaginea de pe ecran. (*image associated with the 1st practice item*)

Vom face mai multe încercări. (*blank screen*)

Ești gata? ("*GATA?*"/"*READY?*" *appeared written in the middle of the screen*)

Hai să începem! (*START appeared written in the middle of the screen*)

English translation
This game is about princesses and superheroes. You will see pictures of Elsa, Anna, Jasmine, Batman, Superman, and Spiderman playing with lots of different animals and people.

Paddington will ask you a question.

You will need to press this key to hear Paddington's question.

Please listen very carefully to the question.

Then you will see a picture on the screen and you will have to answer Paddington's question based on what you see in this picture.

We will do a bit of practice first.

Are you ready?

Let's start!

Appendix B

1. Generalized linear mixed model for the Child group only testing the effect of Age on response accuracy for multiple *who*-questions and multiple *which*-questions (formula: glmer *(Accuracy ˜ WhType + WhOrder + Age + WhType:WhOrder + WhType:Age+ WhOrder:Age + (1 + WhType | ID) + (1 | Item), family = binomial, control = glmerControl(optimizer = "bobyqa"), data)*)

Fixed effects	Estimate	*SE*	*z*-value	*p*-value
(Intercept)	1.927	0.281	6.845	<.0001***
WhType: Who versus Which	3.293	0.411	8.014	<.0001***
WhOrder: SO versus OS	0.083	0.201	0.413	.679
Age	0.054	0.022	2.447	.014*
$\text{WhType}_{\text{Who versus Which}} \times \text{WhOrder}_{\text{SO versus OS}}$	0.771	0.401	1.921	.044*
$\text{WhType}_{\text{Who versus Which}} \times \text{Age}$	0.035	0.028	1.221	.222
$\text{WhOrder}_{\text{SO versus OS}} \times \text{Age}$	0.030	0.014	2.132	.033*

2. Generalized linear mixed model for multiple *who*-questions only for the child group (formula: glmer *(Accuracy ˜ WhOrder + Age + WhOrder:Age + (1 | ID) + (1 | Item), family = binomial, control = glmerControl(optimizer = "bobyqa"), data)*)

Fixed effects	Estimate	*SE*	*z*-value	*p*-value
(Intercept)	3.785	0.536	7.051	<.0001***
WhOrder: SO versus OS	−0.423	0.415	−1.018	.308
Age	0.098	0.039	2.466	.013*
$\text{WhOrder}_{\text{SO versus OS}} \times \text{Age}$	0.008	0.035	0.227	.820

3. Generalized linear mixed model for multiple *which*-questions only for the child group (formula: glmer *(Accuracy ˜ WhOrder + Age + WhOrder:Age + (1 | ID) + (1 | Item), family = binomial, control = glmerControl(optimizer = "bobyqa"), data)*)

Fixed effects	Estimate	*SE*	*z*-value	*p*-value
(Intercept)	0.284	0.162	1.753	.079
WhOrder: SO versus OS	0.471	0.171	2.741	.006**
Age	0.036	0.012	2.949	.003**
$\text{WhOrder}_{\text{SO versus OS}}$ × Age	0.034	0.015	2.235	.025*

Cite this article: Bentea, A. and Marinis, T. (2021). Not all wh-dependencies are created equal: processing of multiple wh-questions in Romanian children and adults. *Applied Psycholinguistics*. https://doi.org/10.1017/S0142716421000059